Everything You Always Wanted To Know About Mathematics*

(*But didn't even know to ask)

A Guided Journey Into the World of Abstract Mathematics and the Writing of Proofs

Brendan W. Sullivan

bwsulliv@andrew.cmu.edu

with Professor John Mackey

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA

May 10, 2013

This work is submitted in partial fulfillment of the requirements for the degree of Doctor of Arts in Mathematical Sciences.

Contents

I Learning to Think Mathematically **11**

1 What Is Mathematics? **13**
1.1 Truths and Proofs . . . 13
1.1.1 Triangle Tangle . . . 14
1.1.2 Prime Time . . . 20
1.1.3 Irrational Irreverence . . . 21
1.2 Exposition Exhibition . . . 22
1.2.1 Simply Symbols . . . 22
1.2.2 Write Right . . . 26
1.2.3 Pick Logic . . . 31
1.2.4 Obvious Obfuscation . . . 37
1.3 Review, Redo, Renew . . . 41
1.3.1 Quick Arithmetic . . . 42
1.3.2 Algebra Abracadabra . . . 43
1.3.3 Polynomnomnomials . . . 49
1.3.4 Let's Talk About Sets . . . 59
1.3.5 Notation Station . . . 60
1.4 Quizzical Puzzicles . . . 61
1.4.1 Funny Money . . . 61
1.4.2 Gauss in the House . . . 65
1.4.3 Some Other Sums . . . 71
1.4.4 Friend Trends . . . 77
1.4.5 The Full Monty Hall . . . 86
1.5 It's Wise To Exercise . . . 92
1.6 Lookahead . . . 98

2 Mathematical Induction **101**
2.1 Introduction . . . 101
2.1.1 Objectives . . . 101
2.1.2 Segue from previous chapter . . . 102
2.1.3 Motivation . . . 102
2.1.4 Goals and Warnings for the Reader . . . 103
2.2 Examples and Discussion . . . 104
2.2.1 Turning Cubes Into Bigger Cubes . . . 104

2.2.2 Lines On The Plane . . . 112
2.2.3 Questions & Exercises . . . 117
2.3 Defining Induction . . . 119
2.3.1 The Domino Analogy . . . 119
2.3.2 Other Analogies . . . 125
2.3.3 Summary . . . 126
2.3.4 Questions & Exercises . . . 127
2.4 Two More (Different) Examples . . . 129
2.4.1 Dominos and Tilings . . . 129
2.4.2 Winning Strategies . . . 133
2.4.3 Questions & Exercises . . . 137
2.5 Applications . . . 137
2.5.1 Recursive Programming . . . 137
2.5.2 The Tower of Hanoi . . . 139
2.5.3 Questions & Exercises . . . 143
2.6 Summary . . . 144
2.7 Chapter Exercises . . . 144
2.8 Lookahead . . . 148

3 Sets **149**
3.1 Introduction . . . 149
3.1.1 Objectives . . . 149
3.1.2 Segue from previous chapter . . . 150
3.1.3 Motivation . . . 150
3.1.4 Goals and Warnings for the Reader . . . 151
3.2 The Idea of a "Set" . . . 151
3.3 Definition and Examples . . . 153
3.3.1 Definition of "Set" . . . 153
3.3.2 Examples . . . 153
3.3.3 How To Define a Set . . . 154
3.3.4 The Empty Set . . . 158
3.3.5 Russell's Paradox . . . 159
3.3.6 Standard Sets and Their Notation . . . 162
3.3.7 Questions & Exercises . . . 163
3.4 Subsets . . . 164
3.4.1 Definition and Examples . . . 164
3.4.2 The Power Set . . . 167
3.4.3 Set Equality . . . 168
3.4.4 The "Bag" Analogy . . . 168
3.4.5 Questions & Exercises . . . 170
3.5 Set Operations . . . 172
3.5.1 Intersection . . . 172
3.5.2 Union . . . 173
3.5.3 Difference . . . 175
3.5.4 Complement . . . 175
3.5.5 Questions & Exercises . . . 176

3.6 Indexed Sets . . . 177
3.6.1 Motivation . . . 177
3.6.2 Indexed Unions and Intersections . . . 181
3.6.3 Examples . . . 181
3.6.4 Partitions . . . 183
3.6.5 Questions & Exercises . . . 184
3.7 Cartesian Products . . . 186
3.7.1 Definition . . . 186
3.7.2 Examples . . . 187
3.7.3 Questions & Exercises . . . 189
3.8 Defining the Natural Numbers . . . 190
3.8.1 Definition . . . 190
3.8.2 Principle of Mathematical Induction . . . 193
3.8.3 Questions & Exercises . . . 193
3.9 Proofs Involving Sets . . . 194
3.9.1 Logic and Rigor: Using Definitions . . . 194
3.9.2 Proving "$\subseteq$" . . . 195
3.9.3 Proving "$=$" . . . 198
3.9.4 Disproving Claims . . . 203
3.9.5 Questions & Exercises . . . 206
3.10 Summary . . . 207
3.11 Chapter Exercises . . . 208
3.12 Lookahead . . . 213

4 Logic 215
4.1 Introduction . . . 215
4.1.1 Objectives . . . 215
4.1.2 Segue from previous chapter . . . 216
4.1.3 Motivation . . . 216
4.1.4 Goals and Warnings for the Reader . . . 216
4.2 Mathematical Statements . . . 217
4.2.1 Definition . . . 218
4.2.2 Examples and Non-examples . . . 219
4.2.3 Variable Propositions . . . 221
4.2.4 Word Order Matters! . . . 224
4.2.5 Questions & Exercises . . . 224
4.3 Quantifiers: Existential and Universal . . . 226
4.3.1 Usage and notation . . . 226
4.3.2 The phrase "such that", and the order of quantifiers . . . 229
4.3.3 "Fixed" Variables and Dependence . . . 230
4.3.4 Specifying a quantification set . . . 232
4.3.5 Questions & Exercises . . . 233
4.4 Logical Negation of Quantified Statements . . . 235
4.4.1 Negation of a universal quantification . . . 235
4.4.2 Negation of an existential quantification . . . 236
4.4.3 Negation of general quantified statements . . . 237

4.4.4 Method Summary . . . 239
4.4.5 The Law of the Excluded Middle . . . 240
4.4.6 Looking Back: Indexed Set Operations and Quantifiers . 241
4.4.7 Questions & Exercises . . . 242
4.5 Logical Connectives . . . 244
4.5.1 And . . . 245
4.5.2 Or . . . 246
4.5.3 Conditional Statements . . . 246
4.5.4 Looking Back: Set Operations and Logical Connectives . 255
4.5.5 Questions & Exercises . . . 256
4.6 Logical Equivalence . . . 258
4.6.1 Definition and Uses . . . 259
4.6.2 Necessary and Sufficient Conditions . . . 263
4.6.3 Proving Logical Equivalences: Associative Laws . . . 264
4.6.4 Proving Logical Equivalences: Distributive Laws . . . 268
4.6.5 Proving Logical Equivalences: De Morgan's Laws (Logic) 269
4.6.6 Using Logical Equivalences: DeMorgan's Laws (Sets) . . . 270
4.6.7 Proving Set Containments via Conditional Statements . . 271
4.6.8 Questions & Exercises . . . 276
4.7 Negation of Any Mathematical Statement . . . 278
4.7.1 Negating Conditional Statements . . . 278
4.7.2 Negating Any Statement . . . 280
4.7.3 Questions & Exercises . . . 282
4.8 Truth Values and Sets . . . 284
4.9 Writing Proofs: Strategies and Examples . . . 286
4.9.1 Proving $\exists$ Claims . . . 287
4.9.2 Proving $\forall$ Claims . . . 291
4.9.3 Proving $\vee$ Claims . . . 293
4.9.4 Proving $\wedge$ Claims . . . 295
4.9.5 Proving $\Longrightarrow$ Claims . . . 297
4.9.6 Proving $\Longleftrightarrow$ Claims . . . 304
4.9.7 Disproving Claims . . . 307
4.9.8 Using assumptions in proofs . . . 309
4.9.9 Questions & Exercises . . . 311
4.10 Summary . . . 312
4.11 Chapter Exercises . . . 313
4.12 Lookahead . . . 319

5 Rigorous Mathematical Induction 321
5.1 Introduction . . . 321
5.1.1 Objectives . . . 321
5.2 Regular Induction . . . 322
5.2.1 Theorem Statement and Proof . . . 322
5.2.2 Using Induction: Proof Template . . . 324
5.2.3 Examples . . . 327
5.2.4 Questions & Exercises . . . 329

5.3 Other Variants of Induction . . . 331
5.3.1 Starting with a Base Case other than $n = 1$ . . . 331
5.3.2 Inducting Backwards . . . 334
5.3.3 Inducting on the Evens/Odds . . . 335
5.3.4 Questions & Exercises . . . 341
5.4 Strong Induction . . . 342
5.4.1 Motivation . . . 342
5.4.2 Theorem Statement and Proof . . . 343
5.4.3 Using Strong Induction: Proof Template . . . 348
5.4.4 Examples . . . 348
5.4.5 Comparing "Regular" and Strong Induction . . . 355
5.4.6 Questions & Exercises . . . 356
5.5 Variants of Strong Induction . . . 357
5.5.1 "Minimal Criminal" Arguments . . . 358
5.5.2 The Well-Ordering Principle of $\mathbb{N}$ . . . 362
5.5.3 Questions & Exercises . . . 364
5.6 Summary . . . 366
5.7 Chapter Exercises . . . 366
5.8 Lookahead . . . 373

II Learning Mathematical Topics 375

6 Relations and Modular Arithmetic 377
6.1 Introduction . . . 377
6.1.1 Objectives . . . 377
6.1.2 Segue from previous chapter . . . 378
6.1.3 Motivation . . . 379
6.1.4 Goals and Warnings for the Reader . . . 379
6.2 Abstract (Binary) Relations . . . 380
6.2.1 Definition . . . 380
6.2.2 Properties of Relations . . . 383
6.2.3 Examples . . . 384
6.2.4 Proving/Disproving Properties of Relations . . . 386
6.2.5 Questions & Exercises . . . 391
6.3 Order Relations . . . 393
6.3.1 Questions & Exercises . . . 398
6.4 Equivalence Relations . . . 399
6.4.1 Definition and Examples . . . 399
6.4.2 Equivalence Classes . . . 402
6.4.3 More Examples . . . 409
6.4.4 Questions & Exercises . . . 412
6.5 Modular Arithmetic . . . 414
6.5.1 Definition and Examples . . . 414
6.5.2 Equivalence Classes modulo n . . . 423
6.5.3 Multiplicative Inverses . . . 433

6.5.4 Some Helpful Theorems . . . 447
6.5.5 Questions & Exercises . . . 455
6.6 Summary . . . 456
6.7 Chapter Exercises . . . 457
6.8 Lookahead . . . 466

7 Functions and Cardinality 467
7.1 Introduction . . . 467
7.1.1 Objectives . . . 467
7.1.2 Segue from previous chapter . . . 468
7.1.3 Motivation . . . 469
7.1.4 Goals and Warnings for the Reader . . . 469
7.2 Definition and Examples . . . 469
7.2.1 Definition . . . 470
7.2.2 Examples . . . 472
7.2.3 Equality of Functions . . . 476
7.2.4 Schematics . . . 480
7.2.5 Questions & Exercises . . . 481
7.3 Images and Pre-images . . . 482
7.3.1 Image: Definition and Examples . . . 482
7.3.2 Proofs about Images . . . 490
7.3.3 Pre-Image: Definition and Examples . . . 493
7.3.4 Proofs about Pre-Images . . . 495
7.3.5 Questions & Exercises . . . 496
7.4 Properties of Functions . . . 497
7.4.1 Surjective (Onto) Functions . . . 497
7.4.2 Injective (1-to-1) Functions . . . 502
7.4.3 Proof Techniques for Jections . . . 506
7.4.4 Bijections . . . 507
7.4.5 Questions & Exercises . . . 509
7.5 Compositions and Inverses . . . 511
7.5.1 Composition of Functions . . . 511
7.5.2 Inverses . . . 516
7.5.3 Bijective $\iff$ Invertible . . . 519
7.5.4 Questions & Exercises . . . 521
7.6 Cardinality . . . 522
7.6.1 Motivation and Definition . . . 522
7.6.2 Finite Sets . . . 528
7.6.3 Countably Infinite Sets . . . 530
7.6.4 Uncountable Sets . . . 549
7.6.5 Questions & Exercises . . . 555
7.7 Summary . . . 557
7.8 Chapter Exercises . . . 558
7.9 Lookahead . . . 566

8 Combinatorics **567**
8.1 Introduction . . . 567
8.1.1 Objectives . . . 567
8.1.2 Segue from previous chapter . . . 568
8.1.3 Motivation . . . 568
8.1.4 Goals and Warnings for the Reader . . . 569
8.2 Basic Counting Principles . . . 570
8.2.1 The Rule of Sum . . . 570
8.2.2 The Rule of Product . . . 574
8.2.3 Fundamental Counting Objects and Formulas . . . 580
8.2.4 Questions & Exercises . . . 588
8.3 Counting Arguments . . . 589
8.3.1 Poker Hands . . . 589
8.3.2 Other Card-Counting Examples . . . 595
8.3.3 Other Counting Objects . . . 604
8.3.4 Questions & Exercises . . . 620
8.4 Counting in Two Ways . . . 623
8.4.1 Method Summary . . . 623
8.4.2 Examples . . . 625
8.4.3 Standard Counting Objects . . . 634
8.4.4 Binomial Theorem . . . 639
8.4.5 Questions & Exercises . . . 641
8.5 Selections with Repetition . . . 643
8.5.1 Motivation . . . 643
8.5.2 Formula . . . 644
8.5.3 Equivalent Forms . . . 645
8.5.4 Examples . . . 647
8.5.5 Questions & Exercises . . . 651
8.6 Pigeonhole Principle . . . 652
8.6.1 Motivation . . . 652
8.6.2 Statement and Proof . . . 653
8.6.3 Examples . . . 654
8.6.4 Questions & Exercises . . . 656
8.7 Inclusion/Exclusion . . . 657
8.7.1 Motivation . . . 657
8.7.2 Statement and Proof . . . 658
8.7.3 Examples . . . 659
8.7.4 Questions & Exercises . . . 662
8.8 Summary . . . 662
8.9 Chapter Exercises . . . 663
8.10 Lookahead . . . 669

A Definitions and Theorems **671**
A.1 Sets . . . 671
A.1.1 Standard Sets . . . 671
A.1.2 Set-Builder Notation . . . 671
A.1.3 Elements and Subsets . . . 672
A.1.4 Power Set . . . 672
A.1.5 Set Equality . . . 673
A.1.6 Set Operations . . . 673
A.1.7 Indexed Set Operations . . . 674
A.1.8 Partition . . . 674
A.2 Logic . . . 675
A.2.1 Statements and Propositions . . . 675
A.2.2 Quantifiers . . . 675
A.2.3 Connectives . . . 676
A.2.4 Logical Negation . . . 677
A.2.5 Proof Strategies . . . 678
A.3 Induction . . . 680
A.3.1 Principle of Specific Mathematical Induction . . . 680
A.3.2 Principle of Strong Mathematical Induction . . . 680
A.3.3 "Minimal Criminal" Argument . . . 681
A.4 Relations . . . 682
A.4.1 Properties of Relations . . . 682
A.4.2 Equivalence Relations . . . 682
A.4.3 Modular Arithmetic . . . 684
A.5 Functions . . . 686
A.5.1 Images and Pre-Images . . . 686
A.5.2 Jections . . . 687
A.5.3 Composition of Functions . . . 687
A.5.4 Inverses . . . 688
A.5.5 Proof Techniques for Functions . . . 688
A.6 Cardinality . . . 692
A.6.1 Definitions . . . 692
A.6.2 Results . . . 692
A.6.3 Standard Catalog of Cardinalities . . . 694
A.7 Combinatorics . . . 695
A.7.1 Definitions . . . 695
A.7.2 Counting Principles . . . 695
A.7.3 Formulas . . . 695
A.7.4 Standard Counting Objects . . . 696
A.7.5 Counting In Two Ways . . . 696
A.7.6 Results . . . 696
A.7.7 Inclusion/Exclusion . . . 697
A.7.8 Pigeonhole Principle . . . 697
A.8 Acronyms . . . 698
A.8.1 General Phrases . . . 698
A.8.2 Induction . . . 698

Part I

Learning to Think Mathematically

Chapter 1

What Is Mathematics?

1.1 Truths and Proofs

How do you know whether something is true or not? Surely, you've been told that the angles of a triangle add to 180°, for example, but how do you *know* for sure? What if you met an alien who had never studied basic geometry? How could you *convince* him/her/it that this fact is true? In a way, this is what mathematics is all about: devising new statements, deciding somehow whether they are true or false, and explaining these findings to other people (or aliens, as the case may be). Unfortunately, it seems like many people think mathematicians spend their days multiplying large numbers together; in actuality, though, mathematics is a far more creative and writing-based discipline than its widely-perceived role as ever-more-complicated arithmetic. One aim of this book is to convince you of this fact, but that's merely a bonus. This book's main goals are to show you what mathematical thinking, problem-solving, and proof-writing are really all about, to show you how to do those things, and to show you how much fun they really are!

As a side note, you might even wonder "What does it mean for something to be true?" A full discussion of this question would delve into philosophy, psychology, and maybe linguistics, and we don't really want to get into that. The main idea in the context of mathematics, though, is that **something is true only if we can *show* it to be true *always***. We know $1 + 1 = 2$ always and forever. It doesn't matter if it's midnight or noon, we can rest assured that equation will hold true. (Have you ever thought about how to show such a fact, though? It's actually quite difficult! A book called the *Principia Mathematica* does this from "first principles" and it takes the authors many, many pages to even get to $1 + 1 = 2$!) This is quite different from, perhaps, other sciences. If we conduct a physical experiment 10 times and the same result occurs, do we know that this will *always* happen? What if we do the experiment a million times? A billion? At what point have we actually *proven* anything? In mathematics, repeated experimentation is not a viable proof! We would need to

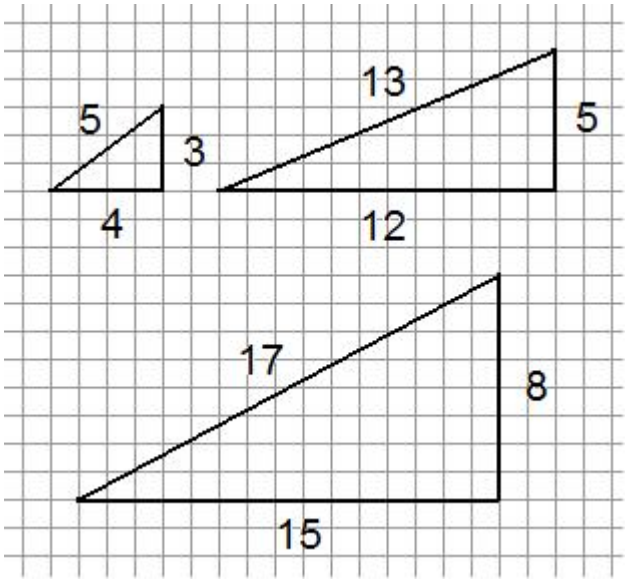
find an argument that shows why such a phenomenon would *always* occur. As an example, there is a famous open problem in mathematics called the *Goldbach Conjecture*. It is unknown, as of now, whether it is true or not, even though it has been verified by computer simulations up until a value of roughly 10^{18}. That's a *huge* number, but it is still not enough to know whether the conjecture is True or False. Do you see the difference? We mathematicians like to *prove* facts, and checking a bunch of values but not *all* of them does *not* constitute a proof.

1.1.1 Triangle Tangle

We've introduced the idea of a **proof** by talking about what we hope proofs to accomplish, and why we would care so much about them. You might wonder, then, how one can *define* a proof. This is actually a difficult idea to address! To approach this idea, we are going to present several different mathematical arguments. We want you to read along with them, and think about whether they are convincing. Do they *prove* something? Are they correct? Are they understandable? How do they make you feel? Think about them on your own and develop some opinions, and then read along with our discussion.

The mathematical arguments we will present here are all about triangles. Specifically, they concern the **Pythagorean Theorem.**

Theorem 1.1.1 (The Pythagorean Theorem)**.** *If a right triangle has base lengths a, b and hypotenuse length c, then these values satisfy $a^2 + b^2 = c^2$.*



How do we know this? It's a very useful fact, one that you've probably used many times in your mathematics classes (and in life, without even realizing).

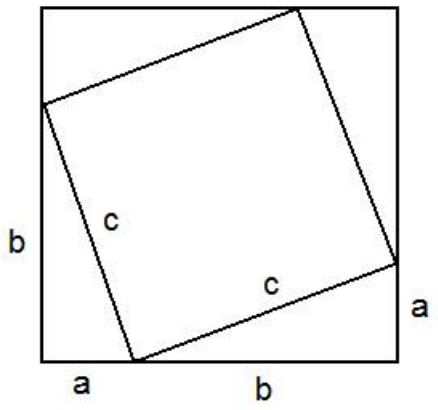
Have you ever wondered why it's true? How would you explain it to a skeptical friend? This is what a **mathematical proof** attempts to accomplish: a clear and concise explanation of a fact. The reasoning behind requiring a proof makes a lot of sense, too, and is twofold: it's a relief to know that what we thought was true is, indeed, true and it's nice to not have to go through the explanation of the fact every time we'd like to use it. After proving the Pythagorean Theorem (satisfactorily), we merely need to refer to the theorem by name whenever a relevant situation arises; we've already done the proof, so there's no need to prove it again.

Now, what exactly constitutes a proof? How do we know that an explanation is sufficiently clear and concise? Answering this question is, in general, rather difficult and is part of the reason why mathematics can be viewed as an art as much as it is a science. We deal with cold, hard facts, yes, but being able to reason with these facts and satisfactorily explain them to others is an art form in itself.

Examples of "Proofs"

Let's look at some sample "proofs" and see whether they work well enough. (We say "proof" for now until we come up with a more precise definition for it, later on.) Here's the first one:

"Proof" 1. Draw a square with side length $a + b$. Inside this square, draw four copies of the right triangle, forming a square with side length c inside the larger square.



The area of the larger square can be computed in two ways: by applying the area formula to the larger square or by adding the area of the smaller square to the area of the four triangles. Thus, it must be true that

$$(a+b)^2 = c^2 + 4 \cdot \frac{ab}{2} = c^2 + 2ab$$

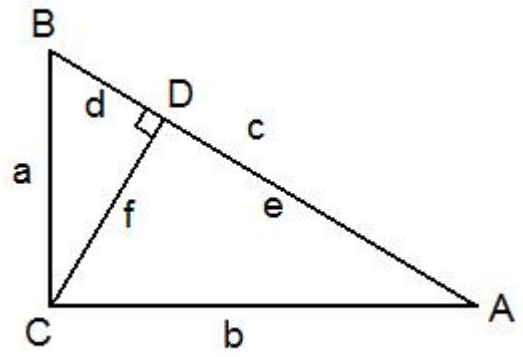
Expanding the expression on the left and canceling a common term on both sides yields

$$a^2 + \cancel{2ab} + b^2 = c^2 + \cancel{2ab}$$

Therefore, $a^2 + b^2 = c^2$ is true. □

Are you convinced? Did each step make sense? Maybe you're not sure yet, so let's look at another "proof" of the theorem.

"Proof" 2. Suppose the Pythagorean Theorem is true and draw the right triangle with the altitude from the vertex corresponding to the right angle. Label the points and lengths as in the diagram below:



Since the Pythagorean Theorem is true, we can apply it to all three of the right triangles in the diagram, namely ABC, BCD, ACD. This tells us (defining $e = c - d$)

$$a^2 = d^2 + f^2$$
$$b^2 = f^2 + e^2$$
$$c^2 = a^2 + b^2$$

Adding the first two equations together and replacing this sum in the third equation, we get

$$c^2 = d^2 + e^2 + 2f^2$$

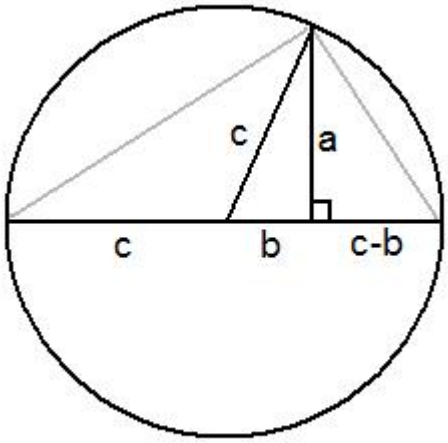
Notice that angles $\angle ABC$ and $\angle ACD$ are equal, because they are both complementary to angle $\angle CAB$, so we know triangles $\triangle CDB$ and $\triangle ADC$ are *similar triangles*. (We are now assuming some familiarity with plane geometry.) This tells us $\frac{e}{f} = \frac{f}{d}$, and thus $f^2 = ed$. We can use this to replace f^2 in the line above and factor, as follows:

$$c^2 = d^2 + e^2 + 2de = (d + e)^2$$

Taking the square root of both sides (and knowing c, d, e are all positive numbers) tells us $c = d + e$, which is true by the definition of the lengths d and e. Therefore, our assumption that the Pythagorean Theorem is true was valid. □

What about this proof? Was it convincing? Was it clear? Let's examine one more "proof" before we decide what a constitutes a "correct" or "good" proof.

"Proof" 3. Observe that



so $\frac{a}{c+b} = \frac{c-b}{a}$ and thus $a^2 + b^2 = c^2$. □

Did that make any sense to you? Finally, here's one last "proof" to consider.

"Proof" 4. The Pythagorean Theorem must be true, otherwise my teachers have been lying to me. □

Discussion

Before reading on, we encourage you to think about these four "proofs" and even discuss them with another student or a friend. What do you think constitutes a "correct" proof? Is clarity and ease of reading important? Does it affect the "correctness" of a proof?

From a historical perspective, mathematical proof-writing has evolved over the years and there is a good, general consensus as to what constitutes a "correct" proof:

- It is important that every *step* in the proof, every logical inference and claim, is *valid*, mathematically speaking.
- It is also important that the proof-writer makes (reasonably) clear why a statement follows from the previous work or from outside knowledge.

What's nice about the *truth* requirement is that mathematics has been built up so that we can read through an argument and verify each claim as True or False. What's difficult to define is *clear* writing. In a way, it is much like Supreme Court Justice Potter Stewart's famous definition of obscenity: "I know it when I see it".

Given these four arguments for comparison, let's assess them for clarity and correctness:

Clarity:

- "Proofs" 1 and 2 are fairly well explained. There are clear statements about what the writer is doing and why. They indicate where any equations come from, and even include some pictures to illustrate their ideas for the reader.

  Notice that "proof 1" does rely on some basic prior knowledge, like the algebraic manipulation of variables and formulae for the area of a triangle and square, but this is fine.

  Likewise, "Proof 2" relies on some understanding of similar triangles and what this means about the lengths of their sides. At least the proof-writers pointed this out, so an interested reader could look up some relevant ideas. If they didn't say this, a reader might be confused and have no idea how to figure out what they're missing!

- "Proof" 3 is very poorly worded! It offers no explanation whatsoever. This makes it quite difficult to determine whether their claims are actually correct. Yes, a picture is included, but there is no indication of *why* they chose to draw a circle around the triangle, or why the stated equations follow from the diagram.

- "Proof" 4 is a grammatically correct English sentence, but it doesn't *explain* anything!

Already, we can see that "Proof" 4 is certainly ont a viable candidate for being a good and proper *proof*. "Proofs" 1 and 2 are still in the running, since they are at least written clearly. "Proof" 3, as it is written now, would probably not be a good candidate; however, maybe it does contain correct ideas that just require better explanations. Perhaps it could be rewritten as a good and proper *proof*.

Let's analayze the logical correctness of these four arguments:

Correctness:

- "Proof" 1 mostly good. The formulae for the areas of the square and triangles are correctly applied, and the algebraic manipulation thereof is correct. But how do we know that the process they described—putting four copies of the given triangle inside a larger square—creates a square with side length c on the inside? They merely say it *does* so without

really saying *why*. Other than this omission, though, this proof is both well-written and correct.

(Can you prove that fact, that the shape inside is actually a square? Just look at its angles: can you show why they are all *right* angles?)

- Unfortunately, "Proof" 2 is completely invalid! Every logical step that it makes does follow from the previous one. For instance, assuming we have the triangles set up this way, we can correctly deduce that ΔCDB and ΔADC are similar triangles. However, why is it that we can *assume* the theorem is True right at the beginning? Isn't that what we are trying to accomplish in the proof, overall? This is a crucial flaw. **Assuming a fact and deducing something True from it does *not* allow us to conclude the original assumption was valid.**

  If this method were valid, we could "prove" just about anything we wanted! Here's an example: What do you think of the following "proof" that $0 = 1$?

  "Proof". Assume $0 = 1$. Then, by the symmetric property of $=$, it is also true that $1 = 0$. Adding these two equations tells us $1 = 1$, which is True. Therefore, $0 = 1$ was a valid assumption, so it must be True. □

  Do you see the similarity between this and "Proof" 2 above? The same sort of flawed reasoning was used: we assumed one fact, did some work to get to something else we know to be True, and then said that the assumed fact must be True, as well.

- Regarding "Proof" 3, most mathematicians would say it is a "bad proof", despite the fact that everything it appears to claim is correct. We say "appears" because, without any words to explain what's going on, we don't actually know what the proof-writer is trying to say! However, we will say that the kernel of a perfectly good proof is contained therein.

  From the diagram, you can show that the stated equation, $\frac{a}{c+b} = \frac{c-b}{a}$, must follow. (Hint: Use similar triangles!) From there, it is a simple manipulation to deduce that $a^2 + b^2 = c^2$.

  Can you write some sentences to go along with the diagram that would turn this into a proper proof?

- Lastly, just about every reasonably logical person (we hope!) would say that "Proof" 4 is not even close to being a proof, however convenient it might be to make such statements.

This discussion shows that "Proof" 1 is actually a good proof. Amongst all four, it is the most clearly-written, and the one that is logically correct. We can refer to it now as a **proof**. "Proof" 2 is outright incorrect, despite how clearly it is presented. "Proof" 3 contains correct ideas, but is not presented clearly. "Proof" 4 is so far from a proof that we don't even want to discuss it.

Question

Before moving on to other topics, we'll leave you with a question: if we give you three positive numbers a, b, c that satisfy $a^2 + b^2 = c^2$, is it necessarily true that there is a right triangle with side lengths a, b and hypotenuse length c? If so, how could you go about constructing it? If not, why not?

1.1.2 Prime Time

While we're on the topic of proofs, let's look at another proof, for a different theorem. As a reminder (or brief introduction), let's talk about *prime numbers*.

Definition, Examples, and Uses

Definition 1.1.2. *A positive integer p that is larger than 1 is called a* **prime** *number if the only positive divisors of p are 1 and p. A non-prime positive integer is called a* **composite** *number.*

Prime numbers have shown to be incredibly important in all branches of mathematics, not just the study of integers and their properties, which is known as **number theory**. One of the most famous **conjectures** (a guess at a theorem that has been neither proven nor disproven thus far) in all of mathematics is the *Riemann Hypothesis*. Its conclusion has been shown to be closely related to the distribution of prime numbers throughout the integers. Many books have been written on this topic. Also, most modern cryptography schemes are based on multiplying huge prime numbers together, relying on the fact that it's quite difficult to undo this process and figure out the two huge prime factors, given their product. So now you know: every time you buy a song on iTunes with your credit card, some computer just multiplied two large prime numbers!

The first few prime numbers are $2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$ (remember, 1 does not fit our definition). How many prime numbers are there? How far apart are they? Is there a pattern? Answering questions like these can be interesting and fun, but also difficult (and sometimes, impossible!). Here, we'll answer one of the questions: are there an infinite number of prime numbers?

Theorem and Proof

Theorem 1.1.3 (Infinitude of the Primes)**.** *There are infinitely-many prime numbers.*

"Proof". Assume instead that there are only finitely-many prime numbers, and list them in ascending order: $p_1, p_2, p_3, \ldots, p_k$, so that p_k is the largest of these prime numbers. Define the new number

$$N = (p_1 \cdot p_2 \cdot p_3 \cdot \cdots \cdot p_k) + 1$$

It must be true that N is divisible by some prime number. However, it cannot be divisible by p_1 or p_2 or $\ldots$ or p_k, because that would leave a remainder of 1,

based on how we defined N. Thus, N is divisible by some other prime number that is *not* in the list.

If N itself is composite (i.e. not prime), then we have found some new prime $p < N$ that is not in the list of *all* primes we presumably had. If N itself is prime, then we have a new prime $N > p_k$, so p_k is not actually the largest prime number. Either way, we have a new prime guaranteed to not be in the given list of k primes. Therefore, there must be infinitely-many prime numbers. □

What do you think of this "proof"? Are you convinced? It feels a little different from the other arguments we've seen so far, doesn't it? Try explaining to a classmate how this one differs from "Proof 1" of the Pythagorean Theorem from the previous section. We will reveal this, though: this "proof" here is actually a fully correct *proof*, sans quotation marks!

1.1.3 Irrational Irreverence

Let's talk about a different type of number, now: **rational** numbers. You might know rational numbers as "fractions" or "quotients" or "ratios".

Definition and Examples

Here is a precise definition of *rational* numbers:

Definition 1.1.4. *A real number r is a* **rational** *number if and only if it can be expressed as a ratio of two integers $r = \frac{a}{b}$, where a and b are both integers (and $b \neq 0$).*

A real number that is not rational is called **irrational**.

Nothing about this definition says that there has to be only one such representation of a rational number; it merely requires that a rational number have at least one such a representation. For instance, 1.5 is a rational number because $1.5 = \frac{3}{2} = \frac{12}{8} = \frac{30}{20}$ and so on. A real number that is not rational is called an **irrational** number, and that's the entire definition: *not* rational, i.e. there is no such representation of the number as a ratio of integers. You may know that $\sqrt{2}$ is an irrational number, but how do you *prove* such a thing? Try it for yourself. We will actually reexamine this question later on (see Example 4.9.4). Other irrational numbers you may know already include e, π, φ and $\sqrt{n}$ where n is any positive integer that is not a perfect square.

Questions

Given this definition of rational/irrational, we might wonder how we can combine irrational numbers to produce a rational number. Try to answer the following questions on your own. If your answer is "yes", try to find an example, and if your answer is "no", try to explain why the desired situation is not possible.

(1) Are there irrational numbers a and b such that $a \cdot b$ is a rational number?

(2) Are there irrational numbers a and b such that $a + b$ is a rational number?

(3) Are there irrational numbers a and b such that a^b is a rational number?

Did you find any examples? It turns out that the answer to all three questions is "yes"! The first two are not too hard to figure out, but the third one is a little trickier.

Here, we'll work through a proof that says the answer to the third is "yes". The interesting part about it, though, is that we won't actually come up with definitive numbers a and b that make a^b a rational number; we'll just narrow it down to two possible choices and show that one of those choices *must* work. Sounds interesting, right? Let's try it.

Proof. We know $\sqrt{2}$ is an irrational number. Consider the number $x = \sqrt{2}^{\sqrt{2}}$. There are two possibilities to consider:

- If x is rational, then we can choose $a = \sqrt{2}$ and $b = \sqrt{2}$ and have our answer.
- However, if x is irrational, then we can choose $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$ because then

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \left(\sqrt{2}\right)^{\sqrt{2}\cdot\sqrt{2}} = \left(\sqrt{2}\right)^2 = 2$$

  and 2 is a rational number.

In either case, we can find irrational numbers a and b such that a^b is a rational number. Thus, such a pair of numbers must exist. □

How do you feel about this proof? Is it convincing? It answers the third question above with a definitive "yes", but it does not tell us *which* pair a, b is actually the correct one, merely that one of the pairs will work. (It turns out that $\sqrt{2}^{\sqrt{2}}$ is also irrational, but that fact takes a lot more work to prove.)

There are plenty of other concrete examples that answer this question, though. Can you come up with any? (Hint: try using the $\log_{10}$ function...)

1.2 Exposition Exhibition

1.2.1 Simply Symbols

Mathematics is a Language

Despite appearances (and some densely-written textbooks), mathematics is not just a collection of symbols that we push around on paper. The English language is based on a fixed group of symbols (the 26 letters of the alphabet plus common punctuation like the period and comma and parenthesis) but we put these symbols together in a specific way, while following some standard and

agreed-upon conventions, to craft meaningful words, phrases, sentences, paragraphs, and so on; in essence, English, like any language, is a way to convey meaning via a collection of symbols and a collection of rules governing those symbols. The same concept applies to the *language of mathematics*: there is a collection of symbols and a set of rules that we apply to those symbols.

One difference is that the collection of symbols we use in mathematics can be rather large, depending on which branch of mathematics currently being discussed. A big part of the structural versatility of mathematics is that we can always create and define new symbols to use. Oftentimes, this is even done to make things shorter and easier to read.

Another main difference between mathematics and other languages is that we choose carefully how to *define* our words and the concepts they represent. Frequently, most of the debates mathematicians have revolve around definitions. This may be surprising to you; perhaps it would make more sense to think that mathematicians debate over proofs and conjectures, or maybe it's a novel idea that mathematicians even debate at all! Choosing the right definitions and terms for a newly-discovered concept is a crucial component of mathematical discovery and exposition since it helps the discoverer/inventor explain his/her ideas to other, interested people. (Without this process, there is no advancement in mathematics, just a bunch of isolated people trying to discover truths on their own.)

The situation is similar with spoken languages, but not as extreme, it seems. For instance, if you said to your friend, "I'm hungry", or "I'm feeling a bit peckish", or "Oh my god, I'm starving", they hear essentially the same message and would respond roughly the same way in each case. In mathematics, though, our definitions are far more precise and don't incorporate the types of nuances that spoken language permits. Of course, there are benefits and disadvantages to both philosophies, but in mathematics we strive for precision whenever possible, so we like our definitions to be exact and unwavering. That said, though, we have control over what those definitions are! This is why debates over definitions are so prevalent in the mathematical world: choosing the right definitions for concepts at hand can make future work with those concepts much easier and more convenient.

Choosing Definitions Properly

As a concrete example, let's return to Definition 1.1.2 of a *prime number* that we saw in the previous subsection. It said:

Definition. *A positive integer p that is larger than 1 is called a* **prime** *number if the only positive divisors of p are 1 and p. A non-prime positive integer is called a* **composite** *number.*

There doesn't seem to be anything questionable about this definition, does there? Perhaps you would have worded it differently or been more concise or used a different variable letter or what have you, but the ultimate message would be the same: a prime number is a certain type of number that has a certain

property. However you choose to write out what that specific type of number is (a positive integer larger than 1) and what that property is (having no positive divisors except itself and 1), you obtain an equivalent definition.

There are some subtle questions behind this definition, though: Why is it that particular type of number? Why is it that particular property—only being divisible by 1 and itself—that we care so much about? What if the definition was slightly different? Would things really change that much? We'll address these questions with another question: What do you think of the following alternative definition of a prime number?

Definition 1.2.1. *An integer p that is less than -1 or greater than 1 is called a* prime *number if the only positive divisors of p are 1 and p.*

Do you notice the subtle difference? All of the numbers that fit the previous definition of "prime" still fit this one, but now so do negative numbers! Specifically, given any number p that is prime under the old definition, $-p$ is now prime under the new definition. Is this a reasonable idea? What's wrong with having negative prime numbers?

How about this third definition of prime numbers?

Definition 1.2.2. *A positive integer p is called a* prime *number if the only positive divisors of p are 1 and p.*

(Remember that 0 is neither positive nor negative, by convention.) Now, the negative numbers are out of bounds, but 1 fits this definition. Is this reasonable? The only positive divisors of 1 are 1 and ... itself, right?

This is where a debate could arise: perhaps you don't mind allowing 1 to be a prime number, but your friend is vehemently against it. Well, without solid reasons either way, there's no way to say that either of you is *wrong*, really; you just made different choices of terminology, and neither of them change the inherent property that the only positive divisors of 1 are 1 and itself. As a similar idea, consider this: whether you call them sandals or thongs or flip-flops, the fact remains that those types of shoes are appropriate footwear at the beach.

With historical hindsight and new desires in mind, though, oftentimes one particular definition is shown to be more appropriate. In the future, we will look at **prime factorizations**, a way of writing every (positive) integer as a product of only prime numbers. For instance, $15 = 3 \cdot 5$ and $12 = 2 \cdot 2 \cdot 3 = 2^2 \cdot 3$ and $142857 = 3^3 \cdot 11 \cdot 13 \cdot 37$ are all prime factorizations.

There is a special property about these factorizations, too: in general, a prime factorization of a positive integer is **unique**! That is, there is one and *only* one way to write a positive integer as a product of prime numbers (since we think of different orderings of the factors as the same thing, so $105 = 3 \cdot 5 \cdot 7$ and $105 = 7 \cdot 3 \cdot 5$ are the same factorization). This is something we will *prove* rigorously using the first definition we gave above. What if we use the second definition, or the third? Is this property of uniqueness still true? Why do you think this uniqueness property is so important? Ultimately, the lesson here is

that definitions should be driven by both logic and usefulness, and this can change over time and stir some debate.

Mathematicians Study Patterns

Another benefit of establishing clear and precise definitions is the knowledge and understanding you gain as a thinker; establishing logical foundations can be helpful in the future. A major aspect of how human beings learn involves identifying patterns through everyday experience and then associating ideas, concepts, words and events with those patterns. Then, one can use those patterns to predict and theorize about abstract ideas, concepts and events.

For instance, it has been studied and shown that human babies initially lack, but develop over time, the concept of *object permanence*. If you show a child a colorful toy that they smile at and enjoy, and then hide it under a cardboard box, the child doesn't quite understand that the toy still exists but is just out of sight. He/she will act as if the object is no longer in existence. At some point, though, we learn that this isn't true and that objects that are outside our realm of vision are still existent. How exactly does this happen? Well, perhaps we recognize the pattern of many such occurrences where an object "disappears" and then we find it again later.

Better examples can be found in the natural sciences, and they illustrate an extra facet of pattern recognition and abstract thinking that is of utmost importantance, particularly in mathematics and the sciences. One can imagine that Neanderthals somehow knew that any time they picked up a rock and held it at arm's length and then let go, the rock would fall to the ground. This probably happened over and over and so they "understood" that this phenomenon is a necessary product of nature. After enough occurrences, it was likely understood that this would always happen, or, at least, any instance in which it didn't happen would cause great confusion and fear. (It is this type of emotional response which might serve to explain how the infrequent but powerful occurrences of, say, volcanic eruptions led ancient civilizations to blame such events on "angry gods").

None of these observations of events brought these prehistoric human beings any closer to understanding *why* the rock would always fall to the ground, or being able to *explain* why it would necessarily happen every time. It would be many milennia before people even began to think to ask why and how this phenomenon occurred, and even longer before Isaac Newton finally proposed a model that sought to explain the behavior of gravity (the name given to this type of phenomenon, eventually). And even now, some say, we still haven't figured out precisely how it works. (Go online and Google "loop quantum gravity" and try to understand that, if you're curious).

It's this abstractive leap in thinking—from observations of a pattern to an epistemological understanding of that pattern—that characterizes a truly inquisitive and intellectual thinker, a true **scientist**, in the best sense of the word. Whom would you consider the better entomologist: the voracious reader who has memorized and can list all of the currently-known species of beetle in the

world, or the laboratory scientist who has examined a variety of species and can take a new specimen and classify it as a beetle or non-beetle? This is somewhat of a leading question, but the main point is this: it is far more beneficial to *understand* a definition and the motivations behind it than it is to simply know a bunch of *instances* that satisfy a certain definition.

This is, arguably, even more important in mathematics. Can you imagine a mathematician who didn't know what a prime number was but could merely list the first 100 prime numbers from memory and was content with that? Of course not! Part of the beauty, versatility, and appeal of the study of mathematics is that we examine patterns and phenomena and then choose how to make the appropriate definitions associated with those patterns. We then use our newfound understanding of those patterns to make rigorously precise predictions about other patterns and phenomena. Thoroughly understanding a definition or concept increases the predictive power, and is far more effective than merely knowing examples of that definition/concept.

1.2.2 Write Right

Another interesting aspect of mathematics is that, as much as it is a language unto itself, we rely on an external language to convey the mathematical thoughts and insights we have. Try rewriting any of the definitions and proofs we've looked at before without using any words. It's tough, isn't it? Accordingly, we want the written language we use to convey mathematical ideas to follow the same types of standards we apply to the mathematical "sentences" we write: we want them to be *precise*, *logical*, and *clear*.

Now, deciding on a precise, logical, and clear definition for each of these three words is a difficult task, in itself. However, we can all agree that it would be ideal for a proof to be:

- **precise:** no individual statement should be untrue or interpretable in multiple ways that would make the truth debatable;
- **logical:** each step should follow from previous steps with proper motivation and explanation; and,
- **clear:** steps should be connected and described with proper English grammar and usage, helping the reader to see what's going on.

Let's examine a few "proofs" that disregard these standards and somehow fail to fit the definition of *proof* that we have so far.

Bad "Proof" #1

First, we have a "proof" that 1=2, so we know there must be something wrong with this one. Can you find the error? Which standard does it violate? Precision, logic, or clarity?

"Proof". Suppose we have two real numbers x and y, and consider the following chain of equalities:

$$\begin{aligned}
x &= y && \\
x^2 &= xy && \text{multiply both sides by } x \\
x^2 - y^2 &= xy - y^2 && \text{subtract } y^2 \text{ from both sides} \\
(x+y)(x-y) &= y(x-y) && \text{factor both sides} \\
x + y &= y && \text{cancel } (x-y) \text{ from both sides} \\
y + y &= y && \text{remembering } x = y, \text{ from the first line} \\
2y &= y && \\
2 &= 1 && \text{divide both sides by } y
\end{aligned}$$

□

The issue here is *precision*. After factoring in line four, it seems convenient and wise to divide by the common factor $(x - y)$ to obtain line five; however, line one tells us that $x = y$ so $x - y = 0$, and **division by zero is not allowed**! Working with the variables x and y was just a way to throw you off the scent and disguise the division by zero. (While we're on the topic, why is division by zero not allowed? Can you think of a reasonable explanation? Think about it in terms of multiplication.)

Bad "Proof" #2

Here's another proof of a similar "fact", namely that $0 = 36$.

"Proof". Consider the equation $x^2 + y^2 = 25$. Rearranging to isolate x tells us

$$x = \sqrt{25 - y^2}$$

and then adding 3 to both sides and squaring yields

$$(x+3)^2 = \left(3 + \sqrt{25 - y^2}\right)^2$$

Notice that $x = -3$ and $y = 4$ is a solution to the original equation, so the final equation should be true, as well. Plugging in these values for x and y tells us

$$0 = (-3+3)^2 = \left(3 + \sqrt{25-16}\right)^2 = (3+3)^2 = 36$$

Therefore, $0 = 36$. □

What happened here? Can you spot the illogical step? Perhaps it would help if we rewrote the steps of the proof using the specific values of the variables

x and y that we chose towards the end:

$$\begin{aligned}(-3)^2 + 4^2 &= 25\\ -3 &= \sqrt{25 - 4^2}\\ (-3+3)^2 &= \left(3 + \sqrt{25-4^2}\right)^2\\ 0 &= 36\end{aligned}$$

It's obvious now, isn't it? There's an issue with applying the square root operation to both sides of an equation, and it's dependent on the fact that $(-x)^2 = x^2$.

When we are looking to solve an equation like $z^2 = x^2$, we have to remember there are two roots of this equation: $z = -x$ and $z = x$. Accordingly, starting from an equation and squaring both sides is a completely logical step (the truth of the resulting equations *follows* from the truth of the original equation), but working the other way is an illogical step (the truth of the squared equation does not *necessarily* tell us that the square-rooted equation is also true). This is an issue with **conditional statements** or **logical implications**, an idea we will discuss in detail later on (in Section 4.5.3). For now, we can summarize this idea with the following line:

$$\text{If } a = b \text{ then } a^2 = b^2, \text{ but if } a^2 = b^2 \text{ then } a = b \text{ or } a = -b.$$

This shows why moving from $x^2 + y^2 = 25$ to $x = \sqrt{25 - y^2}$ in the "proof" above is an illogical step: we are immediately assuming one particular choice for the square root when there are two possible options. What would have happened if we had chosen the negative square root there? Try rewriting the proof with the second step reading $-x = \sqrt{25 - y^2}$, instead, and then use the same values for x and y at the end. What happens? What if you use $x = 3$ and $y = -4$ instead? Or $x = -5$ and $y = 0$? Can you describe how to determine when we should use the positive root x and when we should use the negative root $-x$?

Mathematics Uses the "Inclusive Or"

Since this word just arose, let's mention the use of *or* in the sentence above. When we say "$a = b$ or $a = -b$", we mean that *at least* one of the two statements must be true, possibly both. Now, if both $a \neq 0$ and $b \neq 0$, then only one of the concluding statements can be true; that is, in that context, only one of the roots (positive or negative) will be the correct one and not both. If $b = 0$, though, then both of the concluding statements say the same thing, $a = 0$, so it would be illogical to dictate that *or* means only one of the statements can be true and doesn't allow both of them to be true, simultaneously. In other situations, this distinction makes a more marked difference.

For instance, if you order a sandwich at a restaurant and the waiter asks, "Do you want fries or potato salad on the side?", it is understood that you can choose one of those options, but not both. This is an example of the **exclusive *or*** since

it excludes you from choosing both options. Alternatively, if you forgot to bring a writing implement to class and are looking for any old way to take notes and ask your friend, "Do you have a pencil or pen I can borrow?", it is understood that you really don't care which one of the two options is provided, as long as at least one is available. Maybe your friend has both, and any one of them will do. This is an example of the **inclusive *or***, and this is the interpretation that is assumed in all mathematical examples.

Unclear Arguments

The last two bad "proofs" failed because of issues with precision and logical correctness. The third condition we require of a good proof is that it be *clear*: we want the writing to explain what the proof-writer accomplishes in each step and why that accomplishment is relevant. In other words, we don't want the reader to stop at any point and ask, "What does that sentence mean?" or "Where did that come from?" or similar questions born from confusion. If it helps, think about writing a proof in terms of explaining it to your friend in your class, or the grader who will be reading your homework assignment, or a family member of similar intelligence. Reread your own writing and try to anticipate any questions that might arise or any clarifications that might be asked of you, and then address those issues by rewriting.

There are several ways that a proof could fail this condition and come across as unclear. For one, the words and sentences might fail to properly explain the steps and motivations of the proof, and this could actually be because there are too many words (obscuring the proof by overburdening the reader) or because there are too few words (not giving the reader enough information to work with) or because the words chosen are confusing (not properly explaining the proof). These are issues with the *language* of the proof.

Mathematically, any number of issues could arise, in terms of clarity. Perhaps the proof-writer suddenly introduces a variable without stating what type of number it is (an integer, a real number, etc.) or skips a few steps of arithmetic/algebra or uses new notation without defining what it means first or ... None of these acts is technically wrong or illogical, but they can certainly cause confusion for a reader. Can you think of any other ways that a proof can be unclear? Try to think of a language-based one and a mathematical one.

Bad "Proof" #3

Let's state a simple fact about a polynomial function and then examine a "proof" about that fact. Read the argument carefully and try to pinpoint some sentences or mathematical steps that are *unclear*.

Fact: Consider the polynomial function $f(x) = x^4 - 8x^2 + 16$. This function satisfies $f(x) \geq 0$ for any value of x.

"Proof". No matter what the value of x is that we plug into the function f of x we can write the value that the function puts out by factoring the polynomial,

like this:

$$f(x) = x^4 - 8x^2 + 16 = (x-2)^2(x+2)^2$$

Now, any number z must be less than -2, or greater than 2, or strictly between -2 and 2, or equal to one of them. When $z > 2$ then $z-2$ and $z+2$ are both greater than 0 so $f(z) > 0$. When $z < -2$ then both terms are negative and a negative squared is positive so $f(z) > 0$, too. When $-2 < x < 2$, a similar thing happens, and when $x = 2$ or $x = -2$ one of the terms is 0 so $f = 0$. Therefore, what we were trying to prove has to be true. □

What is there to criticize in this proof? First of all, is it correct? Is it precise? Logical? Clear? Where is it unclear? Try to identify the statements, both linguistic and mathematical, that are even slightly unclear, and try to amend them appropriately. Without pointing out any of the individual errors, we offer below a much better, *clearer* proof of the fact above.

Proof. We begin by factoring the function $f(x)$ by considering it as a quadratic function in the variable x^2

$$f(x) = (x^2)^2 - 8x^2 + 16 = (x^2-4)^2$$

Next, we can factor $x^2 - 4 = (x+2)(x-2)$ and rewrite the original function as

$$f(x) = \big((x+2)(x-2)\big)^2 = (x+2)^2(x-2)^2$$

Now, for any real number x, $(x+2)^2 \geq 0$ and $(x-2)^2 \geq 0$, since a squared quantity is always nonnegative. A product of two nonnegative terms is also nonnegative, so $f(x) = (x+2)^2(x-2)^2 \geq 0$, for any value of x. □

What are the differences between the first "proof" and this second proof? Does your rewritten proof look like this second one, as well?

One of the critiques of the first "proof" is that it does not fully explain the situation where $-2 < x < 2$; rather, it merely says that something "similar" happens and does not actually carry out any of the details. This is a common situation in mathematics (where some steps of a proof are "left to the reader") and it is a convenient technique that can sometimes avoid tedious arithmetic/algebra and make reading a proof easier, faster, and more enjoyable. However, it should be used sparingly and with caution. It is important, as a proof-writer, to make sure that those steps do work, even if you are not going to present them in your proof; you should consider providing the reader with a short summary or hint as to how those steps would actually work. Also, a proof-writer should try not to use this technique on steps that are crucial to the ultimate result of the proof.

In this particular case, the actual steps of factoring are skipped completely and the analysis of the case where $-2 < x < 2$ is only mentioned in passing, yet these are essential components of the proof! It is such a short proof, anyway, that showing these steps does not represent a great sacrifice in brevity or clarity. Again, this brings up the point of proof-writing as an art, as much as a science:

choosing when to leave some of the verification of details to the reader can be tricky. In this particular instance, showing all of the steps is important.

That being said, though, the second proof we showed here is much clearer. Moreover, it completely avoids the case analysis that appears in the first "proof"! There was an issue of clarity with one of the cases in the first "proof", but rather than simply expound the details in the amended version, we opted to scrap that technique altogether and use a shorter, more direct proof. Now, this is not to say that the technique of the first proof is incorrect. Were we to fill in the gaps of the argument of the first "proof", we would obtain a completely correct proof. However, some of the steps in that technique are redundant. Notice that the cases where $-2 < x < 2$ and where $x > 2$ are actually identical, in a sense: the factors satisfy $(x-2)^2 > 0$ and $(x+2)^2 > 0$ in both cases. In fact, this is true of the first case, where $x < -2$, as well! So why separate this argument into three separate cases when the same ultimate observation is applied to all three of them? In this case, it is best to combine them into one (also using the knowledge that when $x = 2$ or $x = -2$, one of the factors is 0). Again, using that expanded technique is certainly not incorrect; rather, it just adds some unnecessary length to the proof.

We mentioned the term "case" and the phrase "case analysis" in the above paragraphs without properly defining or explaining what we mean. For now, we want to postpone a discussion of these terms until we thoroughly discuss logic in Chapter 4. If you're itching for immediate gratification regarding this issue, though, you can skip ahead to Section 1.4.4 and check out the "Hungarian friends" problem, which contains some intricate case analyses.

1.2.3 Pick Logic

We have used the word "logical", and its associated forms, quite frequently, already, without fully explaining what we mean by it. We realize this seems to go against the precision and clarity that we have been so strongly advocating thus far, but we also have to admit that, unfortunately, it is extremely difficult to provide a thorough definition of *logic*.

Games

If you're looking for a decent heuristic understanding of logic, try thinking about it in terms of "logic puzzles" like Sudoku or Kakuro. These puzzles/games are built around very specific rules that are established and agreed upon from the very beginning, and then the solver is presented with a starting board and expected to apply those rules in a rigorous manner until the puzzle is solved. For instance, in Sudoku, remembering the conditions that each digit from 1 through 9 must appear exactly once in every row, column, and 3×3 box allows the solver to systematically place more and more numbers in the grid, continually narrowing down the large number of potential "solutions" to find the unique answer that the starting grid of numbers yields. An important aspect of this solving process is that at no time is it necessary (or wise, at that) to *guess*;

every step should be guided by a rational choice given the current situation and the established rules of the puzzle, and within that framework, the puzzle is guaranteed to be solvable (given enough time, of course).

Mathematical logic is a little different, in some respects, but the essence is the same: there are established rules of how to play the game and every move should be guided by those rules and current knowledge, and nothing else. This is what we mean when we say that writing mathematical proofs should be governed by *logic*: every step, from one truth to another, should follow the agreed-upon rules and only reference those rules or already-proven facts. The "game" or "puzzle" that we're playing in a proof (and in mathematics, in general) is not as clear-cut as a Sudoku puzzle. Even more confusing, though, is the idea that sometimes we start playing an unwinnable game and don't realize it!

This idea of an "unwinnable game" is an astounding, surprising, and downright powerful conclusion of the work of mathematician Kurt Gödel, a 20th century Austrian logician. His *Incompleteness Theorems* address an inherent problem with strong logical systems: there can be True statements that aren't *provable* within that system. We are unable to provide a thoroughly detailed explanation of some terms here (namely, *logical system* and *provable*), but hopefully you see that there is something weird going on here. How could this be possible? If something is True in mathematics, can't we somehow show that it is true? How else would we know that it is true?

Some Mathematical History

To begin to address these natural questions, let's step a little further back in time and discuss the beginnings of logic as a full-fledged branch of mathematics. One thing to keep in mind throughout this discussion is that we can't completely address every topic that comes up, and that this may feel dissatisfying, and we understand that. Part of the beauty of mathematics is that learning about any one topic brings up so many other questions and concepts to think about, and these can be addressed, as well, with more mathematics. Context is important, though, and for the context of this book, we just don't have the time and space to address all of these tangentially-related topics. We are not trying to hide anything from you or sweep some issues under the rug; rather, we're just dealing with the reality of making sure we're not forcing you to read 10,000 pages on the entire history of mathematics just to get our point across!

You will probably study many of the people that we mention below (and the work they did) further along in your mathematical careers. At that point, you'll have a deeper understanding and appreciation for the subject built by hands-on experience with the material, and you'll be better-equipped to tackle the issues therein. For now, we are merely introducing these people out of interest. Mathematics has a rich and interesting history, and it helps to be aware of it! Here, we will try to present a concise yet meaningful interpretation of logic—its history, motivations, and meaning—that fits with the current context.

The mathematicians and philosophers in the mid- to late-19th century who first studied the ideas that would evolve into modern logic were interested in

many of the same issues we are trying to investigate here: How do we know something is True? How can we express that truth? What types of "somethings" can we even declare to be True or not? Breaking down mathematical language to its very roots, these mathematicians studied ways to combine a fixed set of symbols in very specific ways to create more complicated statements, but in the grand scheme of things, these statements were still rather simple. This is not meant to be a knock against their efforts; one must start somewhere, after all, and these people were working from the ground up.

One of the first major efforts was to investigate the foundations of arithmetic, or the study of the **natural numbers** $(1, 2, 3, 4, \dots)$. Much like Euclid sought to study geometry by establishing a short list of accepted truths, or **axioms**, and then derive truths from these given assumptions, Italian mathematician Giuseppe Peano established a set of axioms for the natural numbers, while others approached the topic from a slightly different viewpoint. Meanwhile, this newfound appreciation for being rigorous and decisive about truths and proving those truths led David Hilbert and others to bring up some issues with Euclid's axioms, specifically the parallel postulate.

This work on geometry and arithmetic naturally led into further, intricate study of other areas of mathematics and fervent attepmts to axiomatize fields like analysis of the real numbers. Karl Weierstrass, in studying this topic, produced some mind-blowing examples of functions with strange properties. For instance, try to define a continuous function that is not differentiable anywhere. (If you're unfamiliar with these terms from calculus, don't worry about it; suffice it to say, it's difficult.) Finally, Richard Dedekind was able to establish a rigorous, logical definition of the real numbers, derived entirely from the natural numbers, and not dependent on some vague, physical notion that a continuum of numbers must exist.

Later on, this study branched off slightly into the study of **sets**, or collections of objects. The groundwork for much of this area was laid by Georg Cantor towards the end of the 19th century. He was the first to truly study the theory of infinite sets, establishing the controversial idea that there are different "sizes" of infinity. That is, he showed that some infinite sets are strictly bigger than other infinite sets. This idea was so controversial at the time, that he was hated by many other mathematicians! Nowadays, we realize Cantor was right. (This also gives you a flavor of what we'll discuss later in Section 7.6. Take this as an intriguing example: the set of odd integers and the set of even integers are the same size, sure, but they are both also the same size as the set of *all* integers. However, the set of all real numbers is *strictly* bigger!)

Indeed, some mathematicians were quite shocked by Cantor's discoveries, and even the great Bernhard Riemann thought the development of set theory would be the scourge of mathematics (at first, anyway). This was not the case, though, and it has flourished since then, with many mathematicians working on ways to represent all of mathematics in just the right way and understand the "foundations" of mathematics. In a way, you can think of set theory as the study of the basic objects that all mathematicians are working with, ultimately, in a way similar to the fact that all of chemistry is done by appropriately combining

elements of the periodic table in more and more complicated ways.

A further development from these topics was the study of symbolic logic, which is a bit more concrete than the abstract ideas we've mentioned so far, and whose basic ideas we will be studying frequently in the beginning chapters of this book. This area covers how we can combine mathematical equations and symbols with language-based symbols and connectors to make meaningful mathematical statements that are able to be confirmed as true via a proof. This is an incredibly important component of mathematics, in general, and this book, in particular. Individual viewpoints are certainly more nuanced and specific than this, but, in general, most mathematicians are of the mindset that there are many mathematical truths out there waiting to be discovered and we spend our time learning about the truths we have already uncovered with the hopes of exposing even more of those truths. It's like a giant archaelogical dig, whereby studying the bones and artifacts we've already unearthed will help us to predict what kinds of other treasures we will find and where, which tells us where to look and how to dig once we get there. In a way, logic is that process that is abstracted from the digging by one step: logic is the study of the digging process. It tells us how we can actually take our mathematical knowledge and learn from it and combine it with other knowledge to prove further truths from that.

Now, this is not a precise analogy, mind you, and the study of abstract logic is far more complicated and intricate. For our purposes, in this book, though, this is a sufficiently reasonable way to think of logic. We will learn about some of the first principles and basic operations of symbolic logic and apply this knowledge to our study of writing proofs. It will help us to actually understand what a proof even *is*, it will help guide the construction of proofs that we want to write, it will allow us to critique proofs that may be incorrect, and it will ultimately help us understand just how mathematics works, as a whole.

Applications of Logic: Theoretical Computer Science

One very important application of the ideas and results of logic is in the development and study of computer science, particularly theoretical computer science and computability theory. This particular branch of mathematics was initially motivated by one of David Hilbert's Twenty-Three Problems: this was a list of famously unsolved conjectures in the world of mathematics at the time of their publishing, in 1900. Problem number ten dealt with solving **Diophantine Equations**, which are equations of the form

$$a_1 x_1^{p_1} + a_2 x_2^{p_2} + a_3 x_3^{p_3} + \cdots + a_n x_n^{p_n} = c$$

where $a_1, a_2, \ldots, a_n$ and c are fixed, given constants, $p_1, \ldots, p_n$ are fixed natural numbers, and $x_1, x_2, \ldots, x_n$ are variables that are left to be determined so that they make the equation true.

Given an equation like this, one might wonder whether there are any solutions at all and, if so, just how many there are. Furthermore, if we're given that the fixed constants a_i and c are all rational numbers, we might wonder

whether we can ensure that there is a solution where all of the variables x_i are also rational numbers. Some theoretical results have been established regarding this particular problem, but Hilbert's tenth problem, as stated in 1900, asked whether there was "a process according to which it can be determined in a finite number of operations" whether there is a solution to a given equation where all of the variables x_i are rational numbers. They didn't have a proper notion or definition of this term at the time, but what Hilbert was asking for was an **algorithm** that would take in the values of the constants a_i and c and output True or False depending on whether there exists a solution with the desired property. An important part of his question was that this "process" takes a finite number of steps before outputting an answer.

A student at Cambridge in the United Kingdom by the name of Alan Turing began working on this problem years later by thinking of a physical machine that would be executing the steps required to output an answer to the posed problem. Some subsequent publications of his described his invention, what we now call a *Turing Machine*, which is an interesting theoretical device that could be used to answer some problems in formal logic, but also represents many of the ideas that go into building modern computers. We say it's a *theoretical device* because the nature of its definition ensures that it is not physically feasible to build and operate, but it handles some theoretical problems quite well, including the aforementioned tenth problem of Hilbert. More specifically, this machine gave rise to a proper definition for what we mean when we say that something is computable, or able to be determined in a finite number of steps, and this helped to establish a proper notion of an **algorithm**. It would be unfair of us to discuss these topics without also mentioning Alonzo Church, who was working on similar problems at the same time as Turing. Their names, together, are placed on the *Church-Turing thesis* which relates the work of the Turing machine to the more theoretical, formal logic-based notion of computability.

What Will *We* Do with Logic?

While all of these topics in set theory and logic are inherently interesting and immensely important to mathematics, in general, we simply don't have enough time and space to discuss them in detail. Instead, let's focus a bit more on the notions of logic that we'll be using in writing and critiquing mathematical proofs.

We will consider: (1) what kinds of "things" we can actually state and prove, (2) how we can combine "things" that we know to be true to produce more complex truths, and (3) how we can explain how we arrived at the conclusion that those "things" are, indeed, True. For lack of a better term, we say "things" since we don't yet have a formal definition of **mathematical statement**, which is really the type of "thing" that we will be proving. In essence, a mathematical statement is a combination of symbols and sentences from the languages of mathematics and English (in this book, at least) that can be verified as either True or False, but not both or neither. A proof, then, amounts to arranging a sequence of steps and explanations that use true mathematical statements

and sentences to connect these truths together and yield the desired truth of a specific statement at the end. Our study of logic will deal with just how we can combine those steps and guarantee that our proof leads to the correct assesment of truth at the end.

More specifically, we will examine what a mathematical statement really is and how we can combine them to produce more complicated statements. The words *and* and *or* will be particularly important there, since those two words allow us to combine two mathematical statements together in new and meaningful ways. We will also look at **conditional** mathematical statements, which are statements of the form "If A, then B" or "A implies B". These are really the bread and butter of mathematical statements and a majority of important mathematical theorems are of this form. These statements involve making some *assumptions* or *hypotheses* (contained in the statement A), and using those assumed truths to derive a *conclusion* (contained in the statement B). Look back at the statement of the Pythagorean Theorem in Section 1.1.1 and notice how it is in the form of a conditional statement. (Could it be written another way? Try writing the statement of the theorem in a non-conditional form and think about whether it is an inherently different statement in that form. Find another famous mathematical theorem that is in the form of a conditional statement and try doing the same change of format.)

Another important idea in mathematics, and one that will show up all the time in proof-writing, is the concept of a **variable**. Sometimes we want to talk about a type of mathematical object in generality without assigning it a specific value and this is accomplished by introducing a variable. You have likely seen this happen all the time in your previous study of mathematics, and we've even done it already in this book. Look again at the Pythagorean Theorem statement in Section 1.1.1. What do the letters a, b, c represent? Well, we didn't state it explicitly, but we know that these are positive real numbers that represent the lengths of the three sides of a right triangle. What triangle? We didn't mention a specific one or point to a specific drawing or anything like that, but you knew all along what we were talking about. Moreover, the proofs we examined didn't depend on what those values actually are, merely that they are positive real numbers with certain properties. This is incredibly useful and important and, in a way, it saves time since we don't have to individually consider *all* possible right triangles in the universe (of which there are infinitely-many!) and can reduce the whole idea into one compact statement and proof.

One thing we can do with variables is **quantify** them. This involves making claims about whether a statement is true for *any* potential value of a variable, or maybe for just *one* specific value. For instance, in the Pythagorean Theorem, we couldn't claim that $a^2 + b^2 = c^2$ for any positive real numbers a, b, c; we had to impose extra assumptions on the variables to obtain the result we did. This is an example of **universal** quantification: "For *all* numbers a, b, c with this property and that property, we can guarantee that ..." Similarly, we can quantify **existentially**: "There *exists* a number n with this property."

Can you think of a theorem/fact that we have examined so far that uses existential quantification? Look again at the proof that there are irrational

numbers a and b such that a^b is rational. Notice that this claim we proved is of the existence type: we claimed that *there are* two such numbers with the desired properties, and we then proceeded to show that there must, indeed, be those numbers. Now, the interesting part of that proof was that it was *nonconstructive*; that is, we were able to prove our claim without saying what the numbers a and b actually are, explicitly. We narrowed it down to two choices but never made a claim as to which one is the correct choice, merely that one of the pairs *must* work.

1.2.4 Obvious Obfuscation

As a preview of these logical concepts that we'll be examining in mathematical detail later on, let's take some real-world, language-based examples of these ideas.

Conditional Statements

First, let's investigate **conditional statements**. Mathematical theorems frequently take the form of a conditional statement, but these types of statements also appear in everyday language all the time, sometimes implicitly (which can only add to the confusion). For instance, people talk sometimes about what they would do with their lottery winnings, saying something like

> If I win the lottery, then I will buy a new car.

The idea is that the second part of the statement, after the "then", is dependent on the first part of the statement, which is associated with the "if". When the conditions outlined in the "if" part are satisfied, the actions outlined in the "then" part are guaranteed to take place.

The part of a conditional statement associated with the "if" is known as the **hypothesis** (or sometimes, more formally, the **antecedent**). The part associated with the "then" is known as the **conclusion** (or, more formally, the **consequent**).

Sometimes the conclusion of the conditional is more subtle, or the verb tenses in the sentence are such that it doesn't even include the word "if". Take the following quote from the film *Top Gun*, for example:

> It's classified. I could tell you, but then I'd have to kill you.

The idea here is that the first part, "I could tell you", is a hypothesis in disguise. The sentence *"If I told you, I would have to kill you"* would have the same logical meaning as the actual film quote; however it doesn't convey the same forceful, dramatic connotations. It's quite common to actually not include the word "then" in the conclusion of a conditional statement; while reading the sentence, you might even add the word in your mind without realizing. Take the following lyrics from a song by the band The Barenaked Ladies, say:

> If I had $1,000,000, we wouldn't have to walk to the store.
> If I had $1,000,000, we'd take a limousine 'cause it costs more.

Both lines are conditional statements, but neither includes the word "then"; it is understood to be part of the sentence.

Compare these examples to the following sentence and see what's different:

> I carry an umbrella only if it is raining.

The idea here is that the speaker would hate to carry an umbrella around for no good reason, preferring to make sure it would be of use. Does this sentence have the same meaning as the following, similar sentence?

> If I am carrying an umbrella, it is raining.

In modern language usage, the notion of conditional can be a little fuzzy. The first sentence could be interpreted to mean that sometimes it might be raining but the speaker forgets to bring an umbrella, say. The second sentence is a clear assertion of a conditional statement: seeing me walking around with an umbrella lets you necessarily deduce this is because it's raining. In mathematics, we associate these two sentences and say they have the same logical meaning.

This motivates the meaning of the phrase "only if" and, subsequently, the phrase "if and only if". Consider the following two sentences:

> I will buy a new car if I win the lottery.
> I will buy a new car *only if* I win the lottery.

The first one says that winning the lottery guarantees I will buy a new car, whereas the second one says that the act of buying a new car guarantees that it is because I just won the lottery. If both of these sentences are true, then the events "winning the lottery" and "buying a new car" are equivalent, in a sense, because the occurrence of each one *necessarily guarantees* the occurrence of the other.

Accordingly, mathematical definitions commonly use the phrase **"if and only if"**. For example, we might write "An integer is even if and only if it is divisible by 2." This indicates that knowing a number has that property allows us to call it "even", and knowing a number is even allows us to conclude the divisibility property. (Sometimes, though, a definition will just use *if*, with the *only if* part left unstated but understood. You may have noticed that we did this with the definition of prime numbers in Section 1.1.2.)

Creating More Conditional Statements from Others

Starting with a conditional statement, we can modify it slightly to produce three other conditional statements with the same content but different structure. Continuing to use the "lottery/car" example, let's consider the following four versions of the original sentence:

1. If I win the lottery, then I will buy a new car.
2. If I bought a new car, then I won the lottery.
3. If I don't win the lottery, then I won't buy a new car.

4. If I didn't buy a new car, then I didn't win the lottery.

How do these sentences compare? Do any of them have the same logical meaning as each other? Are all of them True, necessarily, assuming the truth of the first one? We would argue that, in this case, sentence two could be False, even if the first sentence is True. Perhaps I got a hefty raise at work or inherited some money and decided to buy a new car. What about sentences three and four, though? Can they be associated with the others somehow? We will leave this for you to discuss and explore on your own. It might be interesting to ask the same questions of some of the other conditional statements we've looked at and see if your answers are different, too.

One final example of a conditional statement we'll mention comes from a joke by standup comedian Demetri Martin.

> I went into a clothing store and a lady came up to me and said, "If you need anything, I'm Jill." I've never met anyone with a conditional identity before. "What if I don't need anything! Who are you?"

This should give you a flavor for the ways that conditional statements in modern language can be imprecise or subtle, and sometimes open to interpretation. In mathematics, we want these types of statements to be rigorous, well-defined, and unambiguous. This is something we will investigate further later on in Section 4.5.3. For now, though, it might help to think of these types of statements in the rigorous way in which a computer algorithm would interpret an `if...then` statement. When the conditions of the `if` part are satisfied, the subroutine is executed, and they are ignored otherwise. Likewise, a `while` loop is merely a sequence of `if...then` statements condensed into one, concise form.

Quantifiers

Next, let's examine some examples of **quantifiers**. We will use quantifiers when there is an unknown variable meant to be an object drawn from a collection of possible values or representations. For instance, when we quantified the variables a, b, c in the statement of the Pythagorean Theorem, they were drawn from the collection of real numbers that represent the side lengths of right triangles. For a non-mathematical example, consider the following sentence:

> Every person is loved by someone.

What are the variables here? How are they quantified? Be careful because, yes, there are in fact two quantifications in this sentence, one for each of two separate variables. In both cases, the variables represent a member of the collection of all people in the world, and the first variable is quantified universally while the second one is quantified existentially. That may sound confusing, so let's try rewriting the sentence more verbosely:

> For every person x in the world, there exists another person y with the property that person y loves person x.

Do you see how this has the same logical meaning as the first sentence? Surely, this one is unnecessarily wordy and precise for a conversation, but we present it here to show you the underlying variables and quantifiers. The key phrases for the quantifiers are *"for every"* (universal) and *"there exists"* (existential).

The Order of Quantification Matters!

Now, let's look at a similar sentence as the last example:

> Someone is loved by every person.

This sentence is quite similar to the one above; it has all of the same words, even! What did the change in word order do to the logical meaning of the sentence? Well, there are still two variables and two quantifiers, one universal and one existential, but the order in which those quantifiers are applied has been switched. The verbose version of this sentence reads:

> There exists a person x with the property that, for every person y in the world, person y loves person x.

This has a completely different meaning from the first sentence! The first one seemed believable but this one is almost outlandish. This should give you an idea of how important it is to keep the order of quantification straight so that you are actually saying what you mean to be saying.

Nested Quantifiers

The following examples illustrate how our brains can sometimes process quantifiers in language-based sentences fairly quickly and easily, even when the interconnectedness might make it difficult to understand. When quantifiers follow one after the other, we call them *nested*.

The ability to analyze and udnerstand such sentences might depend on the context of the sentence and the message it is trying to convey. If the message makes sense and we believe it, it can be easier to grasp. The best example we know of this phenomenon is embodied by the following quote, attributed to the great presidential orator Abraham Lincoln:

> You can fool some of the people all of the time, and all of the people some of the time, but you can not fool all of the people all of the time.

There are quantifiers all over the place here! We are talking about the collection of all people and the collection of instances in which certain people are fooled, and quantifying on those collections. Try to write this sentence with a few different wordings to see if it can sound any "simpler" or more concise.

Could there be another way to phrase the sentence that would remove some (or all) of the quantifiers without altering the meaning?

Finally, out of personal interest and to inject a bit of humor, we'll mention a similar quote that comes from Bob Dylan's song "Talkin' World War III Blues", from his 1963 album *The Freewheelin' Bob Dylan*:

> Half of the people can be part right all of the time
> Some of the people can be all right part of the time
> But all of the people can't be all right all of the time
> I think Abraham Lincoln said that

We will discuss these topics in greater detail later on, where we will examine their mathematical motivations, meanings, and uses. For now, we can't stress enough how important these issues are in writing proofs. Stringing together a bunch of sentences with no way of knowing how they're connected is not a proof, but a properly-structured series of logical statements and implications is exactly what we're looking for.

1.3 Review, Redo, Renew

Thus far, we have sought to motivate and explain mathematical reasoning and proof-writing from a logical standpoint but, along the way, we have used some mathematical concepts and techniques with which you may or may not be familiar. It is important, of course, to think logically and rationally when doing mathematics, but this is only part of the bigger picture. We have tried to explain how to organize mathematical ideas and structure them in a meaningful way that can convince others of a specific fact, but those ideas must contain some mathematical concepts related to that fact!

For instance, we couldn't have looked at any of the proofs of the Pythagorean Theorem without having a rudimentary understanding of geometry: what a triangle is, some basic properties of triangles and lines and angles, etc. What else did we assume the reader would understand? Many of the steps involved arithmetic, like manipulating multiple equations by multiplying through by the same factor or subtracting two equations, and so on. Those ideas may be second-nature to you now, but at some point you had to learn these things and see why and how they actually worked so that you could safely and appropriately use them in the future.

Look back over some of the other proofs we looked at in the previous sections. What mathematical ideas did we use? Try to write down a few and think about when and how you learned about them. Try to write down some specific facts that we may have used without explicitly saying so and think about why we would do that. Also, try to find a few instances where we made a claim but didn't necessarily fully explain why it must be True, and try to do that. For instance, in "Proof 1" of the Pythagorean Theorem, we drew four identical triangles inside a square and then said that the figure inside would also be a square. Is this really True? How can we be so sure? Try to prove it!

Presumed Knowledge

The main point is that we can't actually write proofs without imbuing them with some meaningful mathematical content. Accordingly, one of the main goals of this book is to share some interesting mathematical facts with you. Sometimes, this involves working with objects you already know about and have seen before (like triangles or prime numbers) and trying to do new things with them. Other times, we may be introducing you to completely new mathematical objects (like equivalence relations or binomial coefficients) and working with those. What we'd like to do now is discuss some mathematical objects and concepts that we will use rather frequently and that you might have seen before. We aren't necessarily assuming that you've seen all of these, but none of these ideas are too hard to learn/relearn quickly, and they will be quite useful throughout the remainder of this book, and the remainder of your mathematical life, as well! We've included a handful of problems for you to work on to give you some practice, both throughout this section and at the end of it.

1.3.1 Quick Arithmetic

We won't be expecting you to multiply six digit numbers in your head or anything like that, but being able to manipulate "small" numbers via addition, subtraction and multiplication is an important skill. Sure, calculators and computing programs can be helpful, but we hope that it isn't necessary to run off to `Maple` or `Mathematica` or your `TI-89` whenever we need to add a couple of four digit numbers, say. Technology provides us with many conveniences in the form of accuracy and time-efficiency but when we rely on these devices too heavily, we diminish our ability to verify those answers we get (in the event of a typo or missed keystroke, for instance) and when we use them too frequently, we may not actually save any time at all!

We encourage you to continually try to perform any arithmetic steps we face either in your head or on a piece of scrap paper. It will be fairly infrequent that any problems/puzzles involve arithmetic with "large" numbers and even when they do, there may be a special trick that can reduce the problem to something easier. For instance, try to work on the following series of problems and see what you notice.

Problem 1.3.1. For each of the following multiplications, try to identify the final digit of the resulting number. If your answer is "0" then try to identify how *many* zeroes are at the end of the resulting number.

1. $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$
2. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 10$
3. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 25$
4. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 100$
5. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 1000$

6. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 10000$

7. $1 \cdot 2 \cdot 3 \cdot \cdots \cdot 10^9$

Try to write down a few sentences that would explain to a friend the procedure you used above. That is, given any number n, explain how to identify the number of zeroes at the end of the number resulting from multiplying $1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$

What did you notice? Did you use your calculator for the first few? Surely that would work, or you could even do the first two or three by hand, but how did that help you later on? How did that help you to explain your procedure? Certainly, you needed to find a more general way of figuring out how to tackle this problem, and resorting to a calculator or computer might help you in some cases, but it won't provide you with any insight into the answer.

If you haven't figured out a general procedure, we'll give you this:

Hint: Think about how many multiples of 2 and how many multiples of 5 appear in the multiplication. Try to pair them together. (Why would you want to do that?)

1.3.2 Algebra Abracadabra

The term **algebra** has a couple of meanings in the mathematical world with some different nuances to each. Usually, the term brings to mind manipulating equations with variables and trying to find a numerical solution for them. This is probably how you handled word problems in a high school algebra class. More generally, though, **abstract algebra** is a branch of mathematics that studies certain mathematical structures that have specific properties, oftentimes having no relation to integers or real numbers.

Much of the groundwork for this field was laid by mathematicians before the 19th century who were seeking roots of polynomial equations where the variables were raised to the third or fourth or even higher powers. For a quadratic equation (containing a variable raised to the second power), you probably remember the formula for finding a root of the equation (i.e. a value for the variable that will make the expression evaluate to zero); this is the famous *Quadratic Formula*.

Did you also know there is a procedure for finding the root of an expression involving a variable that is cubed? Or even raised to the fourth power? Interestingly enough, the mathematicians working on this general problem discovered that there are *no* such procedures possible for higher powers! A lot of the concepts and structures they were working with developed into some inherently interesting mathematics, and people have been studying those objects ever since, eventually branching off and working with the underlying properties of those objects, stripping away the numerical context of finding roots of equations. This is usually what a mathematician means when he/she says "algebra".

In this particular context, though, we will be using "algebra" in the sense that you're likely thinking of it: manipulating multiple equations and variables

to obtain numerical values for the variables that make the expressions evaluate to numbers that satisfy all of the equations involved. There is actually a rich and wonderful theory behind solving systems of linear equations, but this type of in-depth study is better suited for a course on matrix algebra (also called linear algebra). For now, we'll look at a couple of handy tricks and then let you practice them.

Solving Systems of Linear Equations

A system of linear equations is just a collection of equations involving a certain number of variables (all raised to the first power, hence *linear*) multiplied by coefficients and added together, and set equal to some constants. There are specific conditions on the coefficients and constants that guarantee whether or not a solution exists (and whether there are infinitely many or just one, in fact) but we won't get into those specific details. Suffice it to say that the systems of equations we will have to handle in this book will have unique solutions, and this means that the number of equations we have will be the same as the number of variables involved. Knowing that ahead of time, how do we manipulate a system of equations to find that unique solution?

In practice, the most time-efficient way of solving a system depends on the coefficients and constants, as well, and perhaps spotting a particular way of applying the methods we are about to explain. That said, simply following these methods will always work in a short amount of time, anyway, so don't be too concerned with finding the absolute shortest method in any given case.

Method 1: The first method involves a system of two equations and two unknowns. In this case, we can use one of the equations to express one variable in terms of the other, then substitute this into the second equation, yielding one equation and one unknown. From that, we can find a specific value for one variable, and substitute this back into the first equation to find a specific value for the other variable, thereby obtaining the solution we wanted. Let's see this process in action with a particular example. Consider the system of equations below:

$$\begin{aligned} 7x + 4y &= -2 \\ -2x + 3y &= 13 \end{aligned}$$

Following the method we just described, we would rearrange the first equation to write y in terms of x

$$y = \frac{1}{4}(-2 - 7x)$$

then substitute this into the second equation

$$-2x + 3 \cdot \frac{1}{4}(-2 - 7x) = 13$$

and solve that new equation for x:

$$\begin{aligned} -2x - \frac{3}{2} - \frac{21}{4}x &= 13 \\ -\frac{29}{4}x &= \frac{29}{2} \\ x &= -2 \end{aligned}$$

Then, we would use this value in the first equation and solve for y:

$$\begin{aligned} 7 \cdot (-2) + 4y &= -2 \\ 4y &= -2 + 14 = 12 \\ y &= 3 \end{aligned}$$

Therefore, the solution we sought is $(x, y) = (-2, 3)$.

What if we had used the value of x we found in the second equation instead of the first? Well, it would produce the same value of y, but maybe the arithmetic would have been slightly quicker. Or, what if we had done it the other way around, and expressed x in terms of y, solved for y, then went back and solved for x? Again, we would have found the same solution, but perhaps the numbers would work out more "nicely" and save us a few seconds of scratch work. This is what we mean by not worrying about finding the most "efficient" method. Sure, there are multiple ways to approach this system of equations, but they ultimately stem from the same method (substitute and solve) and yield the same solution.

Method 2: An alternative way to handle a system of two equations and two unknowns is to multiply both equations through by particular values and then add them together, choosing those multipliers appropriately so that one of the variables is eliminated. Using the example from above, we could multiply the first equation by 2 and the second equation by 7, making the coefficient of x in both equations equal yet opposite; then, adding the equations reduces the system to one equation and one unknown, just y. Observe:

$$\begin{aligned} 2 \cdot \big(7x + 4y = -2\big) \\ 7 \cdot \big(-2x + 3y = 13\big) \\ 14x + (-14x) + 8y + 21y &= -4 + 91 \\ 29y &= 87 \\ y &= 3 \end{aligned}$$

From there, we can substitute this value into the first or second equation and solve for x.

You can use either of these approaches to handle any system of two equations and two unknowns. Perhaps one would be slightly quicker than the other, depending on the numbers involved, but you won't be saving more than a minute either way, so we encourage you to just choose one and work through it.

Method 3: It can sometimes be convenient to interpret these systems of equations graphically; this is not usually an efficient way of identifying a specific solution to the system, but it can give an indication of whether a solution exists and a rough estimate of the magnitude of the values of the solution.

With two unknowns, we can interpret an equation like $ax + by = c$ in terms of a line in the plane by rearranging: $y = -\frac{a}{b}x + \frac{c}{b}$. This is the line with slope $-\frac{a}{b}$ and y-intercept $\frac{c}{b}$. Given two such equations, we can draw two lines in the plane and locate the point of intersection visually. The (x, y) coordinates of that point are precisely the solution we would find by solving the system of equations as we described above.

This visual method also applies to a system of three equations and three unknowns, but this requires drawing lines in three-dimensional space. This can be difficult to do in practice, but is technically achievable. These same concepts also apply to larger numbers of equations and unknowns, but drawing "lines" in four or more dimensions is impossible for us human beings to visualize!

More than two variables: Reduce! The next part of this method builds upon the first by taking a system of more than two equations (and unknowns) and continually reducing it to smaller systems, eventually obtaining a system of two equations and two unknowns, where we can apply the first part of the method. We will illustrate the method by considering a system of three equations and three unknowns, like the one below:

$$\begin{aligned} 6x - 3y + z &= -1 \\ -3x + 4y - 2z &= 12 \\ 5x + y + 8z &= 6 \end{aligned}$$

The first goal is to eliminate one of the three variables. In essence, this can be done in one of two ways, much like the method for two equations and two unknowns. Let's say we're going to try to eliminate z from the system; we can try to express z in terms of x and y and substitute somehow, or we can multiply some equations and add them together to cancel the coefficients of z. The only difference here is that, whichever option we choose, we need to do it twice. Let's use the first equation to write

$$z = -6x + 3y - 1$$

After substituting this expression for z into both the second and third equations, we will have a system of two equations and two unknowns.

One way to think about this is that we need information from all three equations to ultimately arrive at an answer, and in reducing the system to two equations, we need to somehow retain information from all three of the original equations. The expression we have for z came from the first equation, so we need to subtitute it into the other two to retain all of the information we need.

Compare this to the following sequence of steps: rearrange the first equation to isolate z and substitute this into the second equation, then rearrange the second equation to isolate z and substitute this into the first equation. What

happens? The intuition is that we have somehow "lost" information from the third equation and, yes, we will obtain a system of two equations and two unknowns, but it will have insufficient information to yield a unique solution for x and y. If you actually perform the steps we just described (try doing this to check our work), you obtain the following "system" of two equations after minimal simplification:

$$\begin{aligned} 9x - 2y &= 10 \\ \frac{9}{2}x - y &= 5 \end{aligned}$$

These are really the same equation! Accordingly, we wouldn't actually be able to solve them for unique values of x and y.

Let's return to where we were and substitute the expression for z above into the second and third equations

$$\begin{aligned} -3x + 4y - 2 \cdot (-6x + 3y - 1) &= 12 \\ 5x + y + 8 \cdot (-6x + 3y - 1) &= 6 \end{aligned}$$

and then simplify

$$\begin{aligned} 9x - 2y &= 10 \\ -43x + 25y &= 14 \end{aligned}$$

Applying one of the methods from the first problem will give us the solution $(x, y) = (2, 4)$. Having *both* of these values in hand, we can now return to any one of the original three equations and solve for z; better yet, we can just use the isolated expression for z we found already from the first equation:

$$z = -6x + 3y - 1 = -6 \cdot (2) + 3 \cdot 4 - 1 = -12 + 12 - 1 = -1$$

More than two variables: Reduce another way! Another way to reduce a system from three equations to two equations is related to the "multiply and add" method from before, but we still have to be careful about ensuring that we retain information from all three equations. Using the same system of three equations from above, we might notice that after multiplying the first equation by 8 and the second equation by 4, the coefficient of z in all three equations is either ± 8. This allows us to add/subtract the equations in a convenient way to reduce the system to two equations and two unknowns. Specifically, let's do the multiplication we just mentioned

$$\begin{aligned} 48x - 24y + 8z &= -8 \\ -12x + 16y - 8z &= 48 \\ 5x + y + 8z &= 6 \end{aligned}$$

and then add the first equation to the second

$$\begin{aligned} (48x - 12x) + (-24y + 16y) + (8z - 8z) &= -8 + 48 \\ 36x - 8y &= 40 \end{aligned}$$

and the second equation to the third

$$\begin{aligned}(-12x + 5x) + (16y + y) + (-8z + 8z) &= 48 + 6\\ -7x + 17y &= 54\end{aligned}$$

This produces two equations involving only x and y; furthermore, we have combined information from all three original equations to produce these, so we can be assured that we haven't "lost" anything. Solving this new system

$$\begin{aligned}36x - 8y &= 40\\ -7x + 17y &= 54\end{aligned}$$

via any of the previous methods we discussed produces the solution $(x, y) = (2, 4)$. Substituting these values into any of the three original equations and solving for z produces the ultimate answer we sought.

We could have performed similar steps to eliminate y from the system of three equations, too; for instance, we could add 4 times the first equation to three times the second, and subtract 4 times the third equation from the second. Any of these methods would produce the same ultimate answer, but some of them may shorten the arithmetic steps or involve "nicer" numbers (i.e. fewer fractions, smaller multiplications, what have you). Solving a system with more equations amounts to the same general procedure: multiply the equations and add to eliminate one variable from the system and continue doing this until there are only two equations and two unknowns; then, solve for the values of those two variables and work backwards, substituting those values to solve for the values of the variables that had been eliminated.

Algebra Practice

Problem 1.3.2. Solve the following system of equations for (x, y, z):

$$\begin{aligned}x + y + z &= 15\\ 2x - y + z &= 8\\ x - 2y - z &= -2\end{aligned}$$

Now, solve this similar system for (x, y, z):

$$\begin{aligned}x + y + z &= 15\\ 2x - y + z &= 9\\ x - 2y - z &= -2\end{aligned}$$

Compare the changes in the values of x, y, and z between the two systems.

Which variable changed the most? The least? What is the ratio of these changes?

How large/small can you make this ratio by changing the constant on the right-hand side of the second equation of the system?

Problem 1.3.3. A father, mother and son were sitting in a restaurant eating dinner, when they were approached by another family consisting of a father, mother and son. This second family was struck by their resemblance to the first family, so the second father asked the first, "How old are the three of you? I'm guessing we are all about the same age". The first father happened to be a mathematician and did not feel like giving away his family member's ages so easily, and thus "revealed" them to the others in a tricky way. He said, "Well, our current ages combine to make 72, and I happen to be six times as old as my son. However, in the future when I am just twice his age, our combined age will be twice our current combined age. How old do *you* think we are?"

How old are the three family members?

1.3.3 Polynomnomnomials

Sometimes we will need to work with variables that are squared or cubed or raised to even higher powers. In general, a **polynomial** is the term we use for a function that has one or more variables raised to integer powers, multiplied by coefficients, and added together. Here are some examples of polynomials:

$$x^2 - 7x + 1, \quad 7p^6 + 5p^4 + 3p^2 + 2p, \quad \frac{1}{2}z^2 + 9y^2z - 2y + z^3y^2 - 7z$$

These types of functions are quite common and popular in mathematics, partly due to their convenient properties and partly due to their prevalence in nature. We will see them appear throughout this book. For now, though, let's focus on polynomials that only have *one input variable*.

Roots of Polynomials

Sometimes, we will define a polynomial function in the context of a puzzle and wonder whether there are any values for the input variable that make the ouput value 0. These input values are called **roots** of the polynomial.

One way to identify roots of a polnyomial is to **factor** it into linear terms; that is, we try to express the function as a series of multiplications instead of additions, since we can declare that (at least) one of the factors must be 0 to achieve a 0 value. The motivation behind this technique relies on the following fact:

> **Fact**: If a and b are real numbers and $ab = 0$, then $a = 0$ or $b = 0$ (or possibly both).

Example 1.3.4. Let's see a specific example. Let's try to factor the following polynomial:

$$p(x) = x^2 + 6x + 8$$

(It is common notation to define polynomials as $p(x)$, where p stands for polynomial, x is the input variable, and $p(x)$ is the output value corresponding to the input value x. This doesn't have to be the case, though.)

You might notice that

$$p(x) = x^2 + 6x + 8 = (x + 4) \cdot (x + 2) = (x + 4)(x + 2)$$

(It is also fairly common to drop the $\cdot$ when there are factors separated by parentheses, so we will adopt that convention from here on out, as well.)

The reason this factorization works is because we are applying the distributive property multiple times, in reverse. If we were to expand the factorization we just found, explicitly showing every step, it would look like:

$$\begin{aligned} p(x) &= (x + 4)(x + 2) \\ &= x(x + 2) + 4(x + 2) \\ &= (x^2 + 2x) + (4x + 8) \\ &= x^2 + 2x + 4x + 8 = x^2 + 6x + 8 \end{aligned}$$

All we really did to write down the factorization was to notice that the terms $+4$ and $+2$ have product $+8$, which is the constant term, and they have sum $+6$, which is the coefficient on the x term. Knowing how the subsequent expansion of those factors would work out allows us to write down that factorization without really checking it.

Factoring Quadratics

Let's take what we did with that specific example and try to generalize to any quadratic function. If we want to factor a quadratic polynomial like

$$p(x) = x^2 + bx + c$$

we seek values r and s so that $r \cdot s = c$ and $r + s = b$. Usually, we can do this "by inspection", or by just staring at these two equations and thinking for a minute to come up with the appropriate values. (That's what we did with the last example!)

What do we do if the coefficient of the x^2 is not 1 but some other number a? Well, notice that if we can factor the polynomial $\frac{p(x)}{a} = x^2 + \frac{b}{a}x + \frac{c}{a}$, then we can find a factorization of the original polynomial $p(x)$, as well, by just multiplying by a. This won't affect our ability to find roots of the polynomial (our original goal), because we're assuming $a \neq 0$ (otherwise we didn't really have a quadratic polynomial to begin with and wouldn't need to factor it). Once we've found this factorization, it's easy to identify the roots of $p(x)$; since we want to figure out when $p(x) = 0$, we can just use the factorization and the fact we mentioned above to conclude that

$$\begin{aligned} 0 = p(x) = (x + r)(x + s) \quad &\text{implies} \quad x + r = 0 \text{ or } x + s = 0 \\ &\text{which implies} \quad x = -r \text{ or } x = -s \end{aligned}$$

That is, the roots must be $-r$ and $-s$.

What if we have a polynomial of the form $p(x) = x^2 - a^2$? This particular type of function is known as a **difference of squares**, and has a quick factorzation trick. This is a quadratic polynomial so, following the method from above, we would seek values r, s such that $rs = -a^2$ and $r + s = 0$ (since there is no x term in $p(x)$). The second constraint tells us $r = -s$ and using this in the first constraint tells us $r^2 = a^2$. Thus, using $r = a$ and $s = -a$ achieves the factorization $p(x) = (x - a)(x + a)$ and so the roots are $\pm a$. (Notice that using $r = -a$ and $s = a$ also satisfies both constraints, yet it actually yields the same factorization of $p(x)$.)

Similar tricks can sometimes be applied to polynomials of higher **degree** (recall that "degree" means the highest power of the input variable). For instance, the following polynomial has degree 4

$$p(x) = 4x^4 - x^2 - 3$$

but we can factor it easily if we define $y = x^2$ and write it as a quadratic polynomial

$$p(y) = 4y^2 - y - 3 = (4y + 3)(y - 1)$$

Notice that you can think about the factorizations of the coefficients of the y^2, y, and constant terms to jump right to the factorization we found, or you can follow the division trick we mentioned. Here, we would want to factor $\frac{p(y)}{4} = y^2 - \frac{1}{4}y - \frac{3}{4}$, so we need $rs = -\frac{3}{4}$ and $r + s = -\frac{1}{4}$; using $r = -1$ and $s = +\frac{3}{4}$ works, so we obtain the factorization

$$\frac{p(x)}{4} = (y + (-1))\left(y + \frac{3}{4}\right)$$

which can be simplified as

$$p(x) = 4(y - 1)\left(y + \frac{3}{4}\right) = (y - 1)(4y + 3)$$

which is exactly what we had before.

A Root Yields A Factor

Of course, this trick of identifying roots can work in reverse, too: if we can easily spot a root of a polynomial, that can help us in identifying one of the factors. As an example, look at the cubic polynomial below and see if you can find a root "by inspection"; that is, see if you can find an input value for x that will make $p(x)$ evaluate to zero:

$$p(x) = x^3 - 3x + 2$$

If you haven't spotted it yet, you might want to try plugging in some "easy values", like the first few integers (both positive and negative) to see what happens. If you do so, you'll find that $p(1) = 1 - 3 + 2 = 0$. Accordingly, we

know that a factorization of the polynomial p should include the factor $(x-1)$, since this corresponds to the root $x=1$. Knowing this, we want to divide $p(x)$ by the factor $(x-1)$ so that we can further factor that quotient and identify all of the roots of p.

Polynomial "Division"

How do we divide polynomials, though? We seek another polynomial $q(x)$ so that $p(x) = q(x) \cdot (x-1)$, or in other words, we need to find $\frac{p(x)}{x-1}$. One way to identify such a function is by using the same principles of **long division** that you learned all about in middle school when you were dividing integers. The same concepts apply to polynomial functions! Think back to how long division works, and try working through some basic examples— like $22 \div 7$, say—to refresh your memory about how it works.

Now, let's try to apply those same principles to polynomials. Here's an example of the idea of long division applied to $\frac{x^3-3x+2}{x-1}$:

$$\begin{array}{r|l} & x^2 + x - 2 \\ \hline x-1 & x^3 \quad - 3x + 2 \\ & \underline{-x^3 + x^2} \\ & \quad x^2 - 3x \\ & \quad \underline{-x^2 + x} \\ & \quad\quad -2x + 2 \\ & \quad\quad \underline{2x - 2} \\ & \quad\quad\quad 0 \end{array}$$

In each iteration of the method, we try to find the largest "factor" that can "go into" the larger term. In this case, these are just multiples of powers of x; we identify the largest power of x that can "go into" the current term in question. Since the dividend has x^3 and the divisor has x, we write x^2 above the division line. Then, we multiply $(x-1)$ by x^2, write this below the dividend, and subtract to find the remainder.

We repeat the same process until we have a constant term above the division line (i.e. a multiple of x^0) and see the remainder. Since the remainder here is 0, we know that we have a factorization with no remainder. We can then factor the resultant quadratic by noticing that $r=2$ and $s=-1$ satisfy $r+s=1$ and $rs=-2$, so we can finally write

$$p(x) = (x-1)(x-1)(x+2) = (x-1)^2(x+2)$$

Accordingly, the roots of $p(x)$ are $x=1$ and $x=-2$. For this function, the degree of the polynomial is 3 but the function has only 2 roots. Does this strike you as odd? Can you think of a polynomial of degree 3 that has only 1 root? How about a polynomial of degree 3 with no roots? What about 4 roots, or 5 or more? Are any of these possible? Why or why not? What if we were working with a polynomial of degree 4? Of degree n? What can you say for sure about the number of roots a polynomial has, relative to its degree?

Expanding Factors

Sometimes, when working on a puzzle, we start from a factorization of a polynomial and want to expand the factors completely so we can identify the coefficient of a particular term. How can we quickly and easily multiply polynomials together? In essence, we are trying to apply the distributive law over and over without having to write out all of the steps (although that thorough, step-by-step procedure is guaranteed to work, so if you are unsure of your answer, it is always a good idea to go back and check each step thoroughly).

One particular instance where we can reduce the number of steps involved is when we need to expand a factorization like $(a+b)^n$, where a and b represent any constant or variable and n is an integer. In this specific situation, there is a convenient way to identify the coefficients of the expanded polynomial, and those values come from **Pascal's Triangle**.

This is an arrangement of rows of integers into a triangular shape, where each row corresponds to a particular value of n in such an expansion. The trick to generate Pascal's Triangle is to write the first two rows as all 1s, and the outside "legs" of the triangle as all 1s. In the interior of the triangle, any entry is filled in by finding the sum of the two entries immediately above that entry, to the left and to the right. Try generating the first few rows of the triangle yourself and compare to the one below to make sure you've done the procedure correctly.

$n = 0$: 1

$n = 1$: 1 1

$n = 2$: 1 2 1

$n = 3$: 1 3 3 1

$n = 4$: 1 4 6 4 1

We've written the n values on the left side to indicate the correspondence with the original problem of expanding $(a+b)^n$. In general, any term of the expansion will be some coefficient (taken from the triangle) times $a^k b^{n-k}$ for some value of k between 0 and n; that is to say, in every term of the expansion, the sum of the powers of a and b in that term must be n. The numbers in any given row of the triangle are written in an order corresponding to decreasing powers of a, so that the first 1 is the coefficient of a^n, the next number is the coefficient of $a^{n-1}b$, and so on.

If we were faced with expanding $(a+b)^2$, we would read the $n = 2$ row of Pascal's Triangle and see that the coefficients should be $1, 2, 1$, and that these are the coefficients for a^2, ab, b^2, respectively. Thus,

$$(a+b)^2 = a^2 + 2ab + b^2$$

which we could have also accomplished fairly easily by just expanding by hand. What if we were faced with expanding $\left(x^2+2\right)^4$, say? This isn't done as quickly by hand, so let's see what happens if we use Pascal's Triangle. The $n = 4$ row

tells us the coefficients of $a^4, a^3b, a^2b^2, ab^3, b^4$ are $1, 4, 6, 4, 1$, respectively, where $a = x^2$ and $b = 2$. Thus, we can write

$$\begin{aligned}\left(x^2+2\right)^4 &= 1\cdot\left(x^2\right)^4 + 4\cdot\left(x^2\right)^3\cdot 2 + 6\cdot\left(x^2\right)^2\cdot(2)^2 \\ &\quad + 4\cdot x^2\cdot(2)^3 + 1\cdot(2)^4 \\ &= x^8 + 4\cdot x^6\cdot 2 + 6\cdot x^4\cdot 4 + 4\cdot x^2\cdot 8 + 16 \\ &= x^8 + 8x^6 + 24x^4 + 32x^2 + 16\end{aligned}$$

Try performing this expansion step-by-step and compare, too. There are actually some very interesting properties of Pascal's Triangle that are deeply rooted in some other mathematical concepts, and these properties are particularly useful in the field of **combinatorics**. We will, in fact, examine many of these properties in greater detail later on! For example, you might wonder *why* it is the case that this procedure—adding the two entries above—yields entries that correspond to expanding factors like this. We will prove that it works when we discuss the **Binomial Theorem** and its related ideas! (See Section 8.4.4 if you're curious.)

Completing the Square

There is one more polynomial-related trick we need to mention before deriving an important result. Sometimes, it is useful to rewrite a polynomial as a squared term plus a constant term, so that we can separate the variables and constants in a convenient way. This amounts to adding and subtracting a particular term so that, overall, we have added 0 to the polynomial, but the term is chosen in a way that lets us rewrite the terms of the polynomial conveniently. This process is known as **completing the square**, in the sense that we add a term to create a squared factor, and complete the polynomial by subtracting a corresponding amount.

Let's try this procedure with an example and then attempt to generalize. Start with the following polynomial:

$$p(x) = x^2 + 8x + 9$$

A factorization isn't immediately apparent here, so let's try to complete the square. We want to see a term like $(x + a)^2$, where we know the coefficient of x is 1 since the polynomial has $1 \cdot x^2$. Expanding a term like that gives $x^2 + 2ax + a^2$. Since we need $8x$ to appear, we should use $a = 4$. This expansion gives $x^2 + 8x + 16$, but we really want to see $+9$ as the constant term, so let's add and subtract 7 from the original polynomial:

$$p(x) = x^2 + 8x + 9 + 7 - 7 = (x^2 + 8x + 16) - 7 = (x + 4)^2 - 7$$

Does this look familiar? Precisely, it's a difference of squares, and we know how

to factor that:

$$\begin{aligned} p(x) = x^2 + 8x + 9 &= (x+4)^2 - 7 = (x+4)^2 - \left(\sqrt{7}\right)^2 \\ &= \left(x + 4 + \sqrt{7}\right)\left(x + 4 - \sqrt{7}\right) \end{aligned}$$

Accordingly, the roots of this polynomial are $x = -4 - \sqrt{7}$ and $x = -4 + \sqrt{7}$.

Let's generalize! Suppose we start with a quadratic polynomial of the form

$$p(x) = ax^2 + bx + c$$

and, to complete the square, we want to add and subtract a particular term. How did we find that term before? Well, the expansion of a term like $(rx + s)^2$ yields $r^2x^2 + 2rsx + s^2$, and to match these coefficients with the coefficients of the original polynomial, we see that we need $r^2 = a$, so we should use $r = \sqrt{a}$. (Notice that this requires $a \geq 0$, of course! What should we do if $a < 0$?) Then, to have $2rs = b$, we need $s = \frac{b}{2r} = \frac{b}{2\sqrt{a}}$. Then, when this is expanded we have added on $s^2 = \frac{b^2}{4a}$, so we should subtract that from the polynomial.

These steps are performed below, with some extra algebraic cleanup, of sorts, to make the terms look "nicer":

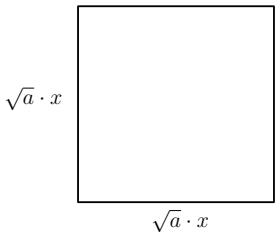
$$\begin{aligned} p(x) = ax^2 + bx + c &= ax^2 + bx + \frac{b^2}{4a} + c - \frac{b^2}{4a} \\ &= \left(\sqrt{a}x + \frac{b}{2\sqrt{a}}\right)^2 + \left(c - \frac{b^2}{4a}\right) \\ &= \left(\sqrt{a} \cdot \left(x + \frac{b}{2a}\right)\right)^2 + \left(c - \frac{b^2}{4a}\right) \\ &= a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right) \end{aligned}$$

This now tells us how to complete the square, given any quadratic polynomial!

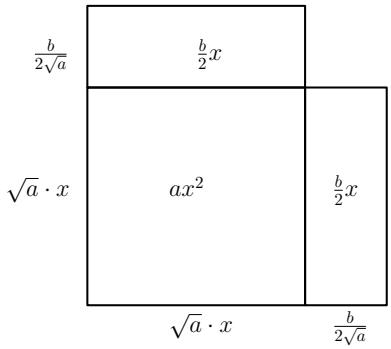
Visualizing Completing the Square

Here's a helpful way to remember how to do this process. It's based on a visual representation of the areas of squares and rectangles.

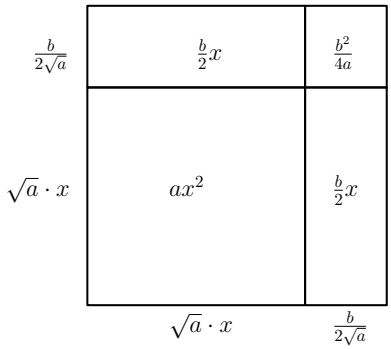
Let's suppose that $a, b > 0$ so that we can geometrically interpret $ax^2 + bx$ as the area of a rectangle. Specifically, let's take the ax^2 term to be the area of a square. This means each side has length $\sqrt{a} \cdot x$:



How might we represent the bx term? We want to build this square into a larger square; this is what completing the square means. So, we should build some rectangles around this square that will help us proceed towards that goal. Let's split the area contributed by the bx term into two rectangles, each with area $\frac{b}{2}x$. Since we must have one side with length $\sqrt{a} \cdot x$, and we want the total area to be $\frac{b}{2}x$, then we see that we need the other length to be $\frac{b}{2\sqrt{a}}$:



What do we need to add to make this a square? We see that there is just a tiny square piece left to fill in the upper-right corner. Each side is length $\frac{b}{2\sqrt{a}}$, so its area—the term we need to add on—is $\frac{b^2}{4a}$.



Look at that! This is the same term we produced above with our algebraic derivation. By adding that on, we were able to factor the terms as a perfect square. We just needed to make sure to subtract it off, as well, so that the net change to the original expression is zero.

This is a helpful trick to keep in mind. It can remind you about both the motivating process for completing the square, as well as how to achieve it. One thing you should ponder, though: Why is it that this visual representation works? We had to assume $a, b > 0$ to be able to draw these diagrams, so why is it that the general formula works no matter what a and b are?

The Quadratic Formula

Let's returning to the question of identifying the roots of a polynomial. Specifically, let's recall the **quadratic formula**. You may have memorized this formula as a way to "solve quadratic equations" but do you know *why* it actually works? Let's try to figure it out! In general, we start with a quadratic polynomial of the form

$$p(x) = ax^2 + bx + c$$

where $a \neq 0$ (otherwise, it's not actually quadratic), and we want to identify the values of x such that $p(x) = 0$. (Did you try to answer our questions above about how many roots this type of polynomial can have? Keep those concepts in mind throughout the following derivation.) We can't hope to factor the polynomial into linear factors too easily, so let's take advantage of the process we used above: completing the square. The benefit of that procedure is that we can set $p(x) = 0$ and rearrange the terms after completing the square to solve for x. Observe:

$$0 = p(x) = ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right)$$

simplifies to:

$$\frac{b^2}{4a} - c = a\left(x + \frac{b}{2a}\right)^2$$

Now, we want to to start "undoing" the processes here to solve for x, and this would require taking the square root of boths sides. But what if $\frac{b^2}{4a} - c < 0$? We couldn't take that square root at all! Or what if $\frac{b^2}{4a} - c = 0$? Is that a problem? Do we have anything to worry about when $\frac{b^2}{4a} - c > 0$? These are issues that are related to the questions we had before about the possible number of roots a polynomial can have. You may have deduced (correctly) that a quadratic polynomial can have *at most* two roots, but here we have uncovered the possibility (and reasons why) that a quadratic polynomial may have one or zero roots!

- In the case where $\frac{b^2}{4a} - c < 0$, then *no* value of x can possibly satisfy the last line in the derivation above. Therefore, there would be no roots of $p(x)$ in the set of real numbers.

- In the case where $\frac{b^2}{4a} - c = 0$, then taking the square root of both sides of the last line above is perfectly valid, but it will produce *exactly one* value

of x:

$$\begin{aligned}\frac{b^2}{4a} - c = 0 &= a\left(x + \frac{b}{2a}\right)^2 \\ 0 &= x + \frac{b}{2a} \\ x &= -\frac{b}{2a}\end{aligned}$$

The remaining case is when $\frac{b^2}{4a} - c > 0$. Here, we can expect *two* roots of $p(x)$ because taking the square root of both sides introduces two possible solutions. In general, when we have a situation like $s^2 = t$, we can say that the only possible solutions are $s = \sqrt{t}$ and $s = -\sqrt{t}$ but we must consider both (we usually write this as $s = \pm\sqrt{t}$). Solving for x in that case yields

$$\begin{aligned}\frac{b^2}{4a} - c &= a\left(x + \frac{b}{2a}\right)^2 \\ \pm\sqrt{\frac{b^2 - 4ac}{4a}} &= \sqrt{a}\left(x + \frac{b}{2a}\right) = \sqrt{a}x + \frac{b}{2\sqrt{a}} \\ -\frac{b}{2\sqrt{a}} \pm \frac{\sqrt{b^2 - 4ac}}{\sqrt{4a}} &= \sqrt{a}x \\ -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{\sqrt{4a^2}} &= x\end{aligned}$$

Now, we need to be careful about the square root observation we made before. In general, $\sqrt{4a^2} = \pm 2a$, but we already know that the fractional term involving that square root already has an associated ± 1 factor, so this factor won't change that. Therefore, we can conclude

$$x = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Voilà, the quadratic formula!

Remember that the final steps of the derivation were carried out under the assumption that $\frac{b^2}{4a} - c > 0$. Does this formula still apply when $\frac{b^2}{4a} - c = 0$? Could we have performed the same steps as we did immediately above while operating under that assumption? Why or why not?

Problems

Problem 1.3.5. Find all possible values of a so that $x - a$ is a factor of $x^2 + 2ax - 3$.

Problem 1.3.6. Find all possible values of b so that $x^3 + b$ is divisible by $x + b$ with no remainder.

Problem 1.3.7. Factor $x^n - 1$ for any natural number n.

Problem 1.3.8. Determine the value of x defined by

$$x = \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}}$$

Hint: try to express the infinitely-nested square roots by using x, itself.

Problem 1.3.9. Use completing the square to prove that the sum of a positive number n and its reciprocal is always greater than or equal to 2, and that the only number that makes the sum equal to 2 is $n = 1$.

Hint: take the sum, add and subtract 2, and rearrange.

Problem 1.3.10. How can we find the roots of a quartic polynomial of the form $ax^4 + bx^2 + c$?

1.3.4 Let's Talk About Sets

We've mentioned some particular types of numbers already, but we want to specifically define the sets of numbers we will be working with in the future. Each of these collections of numbers is represented by a particular letter in the `blackboard bold` font. The **natural numbers** (also known as whole numbers or counting numbers) are so-called because they feel "natural" to say as we start counting objects. We can write

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$$

(There is a more specific and technical definition that we will explain later on.) We use $\mathbb{N}$ to stand for "natural".

Using $\mathbb{N}$, we can define a related collection of numbers: the set of all **integers**, which combines the natural numbers, 0, and the negative natural numbers. We can write

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

The letter $\mathbb{Z}$ comes from the German word *Zahlen*, meaning "number".

From this set, we can define the collection of **rational numbers**. These numbers can be represented as a ratio of integers, but they don't seem to have a natural "listing" like the sets $\mathbb{N}$ and $\mathbb{Z}$, so we can't write this set in the way that we did above. For this, we use a very common set notation, as follows:

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\}$$

We read this as:

> "The set of rational numbers is the set of all numbers of the form $\frac{a}{b}$, where a and b are both integers and b is nonzero."

This conveys the necessary information that a rational number is a fraction, where the numerator and denominator are integers (but the denominator can't

be 0 because division by 0 is disallowed). The reason we use the letter $\mathbb{Q}$ for the rational numbers is because $\mathbb{R}$ was already reserved for the *real* numbers and $\mathbb{Q}$ was the next previous letter available. Also, $\mathbb{Q}$ contains all of the *quotients* of integers, so that makes sense, too!

The **real numbers** $\mathbb{R}$ have a very technical definition that we, unfortunately, cannot delve into completely in this book. (That just goes to show how difficult it is to mathematically define that set!) For now, one way to think of the real numbers is via a **number line**. The real numbers are all numbers that lie on the line, while the numbers of $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ are specific numbers that lie on the line, but they don't comprise the entirety of the line. In a way, $\mathbb{R}$ is the "completion" of $\mathbb{Q}$, in the sense of "filling in the gaps" between rational numbers.

1.3.5 Notation Station

A popular and convenient way of writing sums and products is to use a shortened notation that collects many terms or factors into one common form. For instance, what if we wanted to talk about the sum of the first 500 natural numbers? It would be tedious to write out all 500 terms of the sum, so it is common to write something like $1+2+3+\cdots+499+500$. (We've already used ellipses like this, in fact. Did you understand what we meant?) This is popular and does get the point across, but some mathematicians take offense to the unnecessary use of ellipses in the middle. We put off talking about this issue until now because it's often the case that **notation** can be difficult to learn and comprehend. Rather than bombard you with new symbols right away, we appealed to our intuitive understanding of what "..." accomplish.

Now, that we've brought it up, let's see how to avoid using ellipses. To write the sum we mentioned above, we would use the following notation:

$$1+2+3+\cdots+499+500=\sum_{i=1}^{500} i$$

The large sigma $\sum$ comes from the Greek letter corresponding to S, for "sum", and the **index** i tells us to find the values of the individual terms of the sum. Writing $i=1$ below and 500 above the $\sum$ sign means that we let i assume all of the natural number values between 1 and 500 (inclusive). Using those values, we substitute into the general expression for the term, which is just i, in this case. Accordingly, we find that the terms are $1, 2, 3, \ldots, 500$, as desired. Try to find a few other ways of writing this sum by altering the expression for the general term and/or the values of the index. What if we wanted to find the sum of the first 500 even natural numbers? What about all of the even natural numbers up to (and including) 500? Try to write those sums in the notation style above.

Related to this is the $\prod$ notation. If we wanted to look at the product of the first 500 natural numbers, we would follow the same conventions of identifying

values for the index and the general term:

$$1 \cdot 2 \cdot 3 \cdot \cdots \cdot 499 \cdot 500 = \prod_{i=1}^{500} i$$

The large pi $\prod$ comes from the Greek letter corresponding to P, for "product". Again, try expressing this in a different way by changing the general term and/or index values. What if we wanted to find the product of the first 500 *even* natural numbers? What about all of the even natural numbers *up to* (and including) 500? Try to write those products in the notation style above.

Problems

Problem 1.3.11. Write an English sentence that describes what the following equation means:

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

Problem 1.3.12. Express, in appropriate notation, the sum and product of the first n powers of 2, starting with $2^0 = 1$. Can you prove a formula for the sum? The product?

Problem 1.3.13. Consider the sum of all the odd numbers between 17 and 33 (inclusive). Write this sum in summation notation where the index starts at 0. Now try writing it where the index starts at 1. Now try writing it where the index starts at 8, and again with 9. Which of these feels "more natural"? Why?

1.4 Quizzical Puzzicles

Let's put into action some of the principles we have discussed so far. Specifically, let's examine some interesting mathematical puzzles and explain how to go about solving them. None of these involves knowing anything beyond basic algebra and arithmetic, but this does not mean they are "basic" or "easy", since they all involve critical thinking skills and keen insight to solve and understand. Along the way, we will be employing some logical ideas we have brought up already. We might have to work with polynomial functions, or solve some equations algebraically. We might have to think careful about order and flow of our arguments, making sure everything follows from previous knowledge or deductions. Overall, we should also be thinking about what constitutes a good and valid *proof* of the facts we discover!

1.4.1 Funny Money

Problem Statement

This classic puzzle is contained in a story about some friends paying for a shared hotel room:

Three friends are on a road trip and stop at a hotel late one night looking for a room to catch up on some rest. The clerk on duty says that there is only one room vacant for that night and it costs \$30 for the three of them, if they want to squeeze in together. The friends decide they are desperate for sleep and agree to split the room amongst them, each placing a \$10 bill on the counter to pay up front. The clerk thanks them, hands them the key, and they head off to grab their bags from the car. Meanwhile, another clerk shows up to start his shift and realizes that the previous clerk had made an error and overcharged the three friends for their room: it should have only been \$25. He takes a \$5 bill from the register, hands it to the bellhop on duty and says, "Bring this to Room #29. The guests there were overcharged." The bellhop nods and heads off to their room. When the three friends answer the door, they are surprised and happy to discover they earned a refund. To split the money fairly, one friend makes change with five \$1 bills and then each friend takes \$1, giving the remaining \$2 to the bellhop as a friendly tip. The bellhop thanks them kindly and heads back to work.

Now, each of the three friends contributed \$9 to the room, plus a \$2 tip, making a total of \$29. But they originally gave the clerk \$30... What happened to the missing dollar?!

Think carefully about this before turning the page and reading our solution.

Solution: Keeping Careful Track of the Money

Did you figure it out? Did you realize there is really nothing "missing"? This puzzle is intended to confuse the reader and mislead them into searching for something that isn't really there. The numbers involved are chosen so that the "missing sum" of \$1 is so small that the reader believes that something mysterious happened, but careful and logical analysis of the events should lead you to realize that the question at the end is not really a fair one; it is based on a misinterpretation of the situation, and trying to ignore its reasoning is the key to figuring out the solution to this puzzle. When the numbers are changed greatly so that the final discrepancy is of a much larger value, the reader no longer has that emotional investment to seek out that "missing dollar".

First, let's analyze what actually happened in this particular case. The key is careful interpretation of where the money actually goes. It helps to forget about the individual people involved and think of two distinct entities: the group of friends, who we'll call F, and the clerk/bellhop combo from the hotel, who we'll call H. Now, let's replay the steps of the story and describe where the money comes from and goes to at each step:

(1) F arrives and gives \$30 to H (original cost of room)

(2) H gives \$5 back to F (refund for overcharge)

(3) F gives \$2 back to H (tip to bellhop)

(4) Net change: F gave \$30-\$5+\$2=\$27 to H

It makes more sense now, doesn't it? The refund was \$5, so the room actually cost \$25 and it doesn't make sense to say that the three friends each paid \$9 for it, plus the tip to the bellhop. That \$27 contribution from the three of them *includes* the tip. The question after the story implies that the we should be *adding* the tip to the friends' contribution but, really, it is part of their contribution. By grouping the friends together and the bellhop/clerk together, we can actually track down how the money changes hands.

Generalizing: Changing The Numbers

Let's change the problem in the way that we mentioned above; specifically, let's try to change the numbers of the problem to remove that emotional attachment to the "missing dollar" and make the discrepancy much larger. To begin, we will define some variables to represent the dollar amounts used in each of the steps outlined above. We could try to approach this problem by "testing out" specific values for these dollar amounts and seeing what happens but it will be more efficient to essentially "try everything at once" by introducing variables and substituting specific values for them later on.

We will let $3n$ represent the original cost of the hotel room (the amount paid by the three friends when they first arrived), for some value of n. We choose this because we want the cost to be evenly split by the friends. Next, we want

to define a variable to represent the refund they receive. Knowing that they will want to split this amount evenly amongst the three of them and have some leftover to tip the bellhop, let's say that the refund is of the form $3r + 2$. The variable r represents how much each friend individually receives back from the hotel, and the 2 represents the tip for the bellhop. Now, let's restate the puzzle with these variables instead of the original values.

> Three friends are on a road trip and stop at a hotel late one night looking for a room to catch up on some rest. The clerk on duty says that there is only one room vacant for that night and it costs $\$3n$ for the three of them, if they want to squeeze in together. The friends decide they are desparate for sleep and agree to split the room amongst them, each placing $\$n$ on the counter to pay up front. The clerk thanks them, hands them the key, and they head off to grab their bags from the car. Meanwhile, another clerk shows up to start his shift and realizes that the previous clerk had made an error and overcharged the three friends for their room: it should have only been $\$3n - (3r + 2)$. He takes $\$3r + 2$ from the register, hands it to the bellhop on duty and says, "Bring this to Room #29. The guests there were overcharged." The bellhop nods and heads off to their room. When the three friends answer the door, they are surprised and happy to discover they earned a refund. To split the money fairly, each friend takes $\$r$, giving the remaining \$2 to the bellhop as a friendly tip. The bellhop thanks them kindly and heads back to work.
>
> Now, each of the three friends contributed $\$n - r$ to the room, plus a \$2 tip, making a total of $\$3(n - r) + 2$. But they originally gave the clerk $\$3n \ldots$ What happened to the missing $\$3n - [3(n - r) + 2] = \$3r - 2$?!

Do you see what happened now? The discrepancy occurs, as we explained before, because the question considers *adding* the \$2 tip to the refunded cost of the room and comparing that to the original cost of $\$3n$. The actual comparison should be between the refunded contribution of the friends, which is $\$3(n-r) = \$3n - 3r$, and the sum of the refunded cost of the room and the tip, which is $\$3n - (3r + 2) + 2 = \$3n - 3r$. No discrepancy there!

Generalizing: Questions For You

In the original statement of the puzzle, the values were $n = 10$ and $r = 1$, so that the "missing amount" was magically $\$3r - 2 = \$1$. If we had chosen larger values—say $n = 100$ and $r = 10$—then the \$300 room actually should have cost \$268, the bellhop would bring the friends \$32, they would each take \$10, he would keep \$2, and the discrepancy becomes \$28. Would anyone actually believe that \$28 went missing in those transactions? What if we use even larger values of n and r? How large can you make the discrepancy? How small? Given

a desired discrepancy, in dollars, can you find values for n and r that achieve that value? How many ways are there to do that?

Lessons From This Puzzle

Logic and rational thinking are important when solving a puzzle because it is sometimes easy to be misled by emotions. Had we stated the puzzle originally as the "missing \$28 problem", would you have reacted the same way? Would you have been as temporarily confused before trying to backtrack and discover what really happened?

1.4.2 Gauss in the House

Problem Statement

There is a popular anecdote among mathematicians that may or may not be apocryphal, but some of us would like to believe that it is true, because it features one of the greatest mathematicians/physicists of all time, Carl Friedrich Gauss. He worked in the late 1700s and early- to mid-1800s and proved some fundamental and powerful results in a broad range of areas. He studied number theory, and complex analysis, and optics, and geometry, and astronomy, and so much more! Read the story below, think about what you would have done in that situation—as a young child and now—and then read on for a discussion.

> It was early in the morning in an elementary school classroom, and the students were acting noisy and rowdy, much to the dismay of the teacher, who was feeling quite sick and tired, literally, and quite sick and tired of their behavior. He needed a way to occupy them for a while so that he could relax at his desk and recuperate. He bellowed out to the room and told them to take out their slates and chalk. After asking a few more times, everyone had obliged. He then told them to add together all of the numbers from 1 to 100, and that the first person to do so would earn the privelege of being the teacher's helper for the day. He returned to his desk and sat down, relieved that they would be occupied for quite some time performing large sums. After only a minute, though, one boy walked up to the desk and showed the teacher his slate with the answer. The teacher was astounded and had to spend a few minutes performing the arithmetic himself to check the answer, but in the end, the little boy was correct, and he had accomplished this feat so quickly. How did he do it?

Think carefully about this before turning the page and reading our solution. Remember, this story "happened" in the days before calculators, so you should not be using anything more than your brain and a pencil and paper.

Solution: Reducing Computations

Perhaps you figured this one out. There are actually a number of ways to approach this problem that are slightly different, but they mostly amount to the same insight: trying to reduce the number of computations required.

To naively go through and add each of the 100 numbers to the previously obtained sum would require 99 additions, with ever larger numbers involved. Certainly, the trick here is not to just do these additions faster than the others, it is to be more efficient with the computations that are required. Remember that multiplication can be viewed as repeated addition of one number with itself, so perhaps we can reduce all of these additions to one multiplication, provided we find the right number to add to itself over and over.

Another important fact to remember is that addition is **associative** and **commutative**, meaning we can perform the additions in any order and be assured that we obtain the same answer. Specifically, we can add all the numbers from 100 down to 1 and get the same result for the sum, call it S. Let's write down this fact in a convenient way here:

$$\begin{array}{ccccccccccccccc} 1 & + & 2 & + & 3 & + & \cdots & + & 98 & + & 99 & + & 100 & = & S \\ 100 & + & 99 & + & 98 & + & \cdots & + & 3 & + & 2 & + & 1 & = & S \\ \hline 101 & + & 101 & + & 101 & + & \cdots & + & 101 & + & 101 & + & 101 & = & 2S \end{array}$$

Notice that we have written down the desired sum in two different ways, added those two sums entry by entry, and obtained an expression for $2S$, twice the desired sum. That new expression can be written as a multiplication because there are 100 terms, each of which is the number 101. Thus,

$$2S = 101 \cdot 100 \quad \text{and therefore,} \quad S = 101 \cdot 50 = 5050$$

This is much faster than performing 99 additions, and in fact, if we think carefully, we may be able to do the entire process in our heads!

Alternate Solution: Pairing Terms

Now, a very similar way of seeing this problem is to skip adding the two lines we wrote above and just pair off the numbers in the original sum, as follows:

$$\begin{aligned} S &= 1 + 2 + 3 + \cdots + 98 + 99 + 100 \\ &= (1 + 100) + (2 + 99) + (3 + 98) + \cdots + (49 + 52) + (50 + 51) \\ &= 101 + 101 + \cdots + 101 = 50 \cdot 101 = 5050 \end{aligned}$$

This approach is essentially equivalent to the one we described above; it still takes advantage of the associative property of addition to convert the sum into a multiplication, it just skips over the intermediate steps where we found an expression for $2S$ and then divided by two.

Generalizing: Even n

What if the teacher had asked his students to add the numbers from 1 to 1000? Would they have protested? Would Gauss have been able to find the answer just as quickly? What would you do? We're not sure about the first two questions, but we think that you could handle this sum just as easily. The only thing different here is that the number of pairs we create will be 500 (instead of 50), and each of those pairs will sum to 1001 (instead of 101), so the result will be

$$1 + 2 + 3 + \cdots + 998 + 999 + 1000 = 1001 \cdot 500 = 500500$$

Does it look like there's a pattern there? Do you think you could say what the sum of all of the numbers from 1 to 1 million is right away without doing the multiplication?

Generalizing: Odd n

What if the teacher had asked for the sum of the first 99 numbers instead? Would the pairing process still work? Let's see:

$$\begin{aligned} S &= 1 + 2 + 3 + \cdots + 97 + 98 + 99 \\ &= (1 + 99) + (2 + 98) + (3 + 97) + \cdots + (48 + 52) + (49 + 51) + 50 \\ &= (49 \cdot 100) + 50 = 4950 \end{aligned}$$

Notice that we had an *odd* number of terms in total, so we couldn't pair off every number, and had to add 50 to the result of the multiplication. Could we have paired the numbers in a different way?

$$\begin{aligned} S &= 1 + 2 + 3 + \cdots + 97 + 98 + 99 \\ &= (1 + 98) + (2 + 97) + (3 + 96) + \cdots + (48 + 51) + (49 + 50) + 99 \\ &= (49 \cdot 99) + 99 = 50 \cdot 99 = 4950 \end{aligned}$$

This *seems* more similar to the result of the original puzzle, because we ultimately performed *one* multiplication. This may seem like a strange coincidence now, but try to follow the steps above with some other odd sums. What is the sum of the first 7 integers? The first 29? The first 999? The first 999999?

Generalizing: Any n

Let's step back from the individual cases that we have examined here and try to solve the problem in a more general sense. Let's pretened that the teacher had presented the students with the following problem:

> Find a formula for the sum of the first n numbers. I want a *specific* formula so that if someone tells me what n is, I can find an answer quickly by plugging in that particular value.

The caveat in the second sentence rules out a solution of the form given by our investigations above. We have already come up with some simple *algorithms* for finding a solution to this problem, but we have now been asked to find a *forumla* that will produce a solution. How do we begin to approach this? Well, based on some observations we made above, it would make sense to tackle this puzzle by handling the cases where n is even and where n is odd separately. We saw that the pairing worked out slightly differently in those cases, so let's investigate one and then the other. In each case, we are looking for a formula for $S(n)$, the sum represented by $1 + 2 + 3 + \cdots + (n-2) + (n-1) + n$. We are using this new notation $S(n)$ to indicate that the sum *depends* on that particular value of n.

If n is even, we know that we can pair off every number and have no terms leftover:

$$\begin{aligned} S(n) &= 1 + 2 + 3 + \cdots + \left(\frac{n}{2} - 1\right) + \frac{n}{2} + \left(\frac{n}{2} + 1\right) + \cdots \\ &\quad + (n-2) + (n-1) + n \\ &= (1+n) + (2 + (n-1)) + (3 + (n-2)) + \cdots \\ &\quad + \left(\left(\frac{n}{2} - 1\right) + \left(\frac{n}{2} + 2\right)\right) + \left(\left(\frac{n}{2}\right) + \left(\frac{n}{2} + 1\right)\right) \\ &= (n+1) \cdot \frac{n}{2} = \frac{n^2 + n}{2} \end{aligned}$$

Try this formula with some of the even values of n we examined above (like 100, 1000, 1000000, etc.) It works, doesn't it? Note that the reason we can write terms involving $\frac{n}{2}$ and be assured they are part of the sum is that n is *even*, so $\frac{n}{2}$ is also a whole number.

Okay, now what happens if n is odd? We know that we won't be able to pair off every number, so we need to be clever about what we do here. Remember our approach with summing the first 99 numbers? By leaving off the last term of the sum, we could pair off all of the other terms with no leftover, and futhermore, each of those pairs summed to the *same value* as that last number. Let's try using that approach here:

$$\begin{aligned} S(n) &= 1 + 2 + 3 + \cdots + \left(\frac{n-1}{2} - 1\right) + \frac{n-1}{2} + \left(\frac{n-1}{2} + 1\right) + \cdots \\ &\quad + (n-2) + (n-1) + n \\ &= (1 + (n-1)) + (2 + (n-2)) + \cdots + \left(\left(\frac{n-1}{2}\right) + \left(\frac{n-1}{2} + 1\right)\right) + n \\ &= n + n + \cdots + \left(\frac{2n-2}{2} + 1\right) + n = (n + n + \cdots + n) + n \end{aligned}$$

This has shown that each of the pairs of terms sums to n, the final number that we removed before the pairing process. Now, let's think carefully about how *many* pairs we had. Notice that we can number them by looking at the first number of the pair: the first pair was $(1, n-1)$, the second pair was $(2, n-2)$, and so on, and the first number in the last pair was $\frac{n-1}{2}$. Therefore, we had

exactly that many pairs: $\frac{n-1}{2}$. (Remember that n is odd, so we can rest assured that $n-1$ is even, so $\frac{n-1}{2}$ is a whole number. We haven't been mentioning that every time, so be sure to go back over what we've done so far and convince yourself that every step and every term we write is valid.) To those pairs, we tacked on a final number, n, so we can write the multiplication for the sum as follows:

$$S(n) = \left(\frac{n-1}{2} + 1\right) \cdot n = \left(\frac{n-1}{2} + \frac{2}{2}\right) \cdot n = \frac{n+1}{2} \cdot n = \frac{n^2+n}{2}$$

Wow, this is the exact same formula we found in the case where n is even! Did this surprise you? It's not obvious at all that we should end up with the same formula, even with the similarity of the approaches to the problem. What does this suggest to you? A mathematician would see such a "coincidence" and wonder whether there is a much *simpler* and *direct* route to this result; that is, is there a way we could approach this puzzle that would answer *both* odd and even cases simulatenously? Since we obtained the same answer, there might be a way to do it. Think about this for a minute before reading on.

Generalizing: Any n, *without* separate cases

It turns out that we already hinted at this other method in our previous discussion of this puzzle. Remember when we wrote the sum forwards on one line and backwards on another line and added them together? Well, when we treated the odd/even cases here, we decided to avoid that method because it seemed to add a couple of extra steps; the "pairing terms" process seemed slightly quicker so we followed that method. What if we went back and reexamined the "add the sum twice" method? We'd find something like this:

$$\begin{array}{ccccccccccc} 1 & + & 2 & + & \cdots & + & (n-1) & + & n & = & S(n) \\ n & + & (n-1) & + & \cdots & + & 2 & + & 1 & = & S(n) \\ \hline (n+1) & + & (n+1) & + & \cdots & + & (n+1) & + & (n+1) & = & 2S(n) \end{array}$$

In this case, we have n terms in the sum on the third line, and each term is $(n+1)$. Thus,

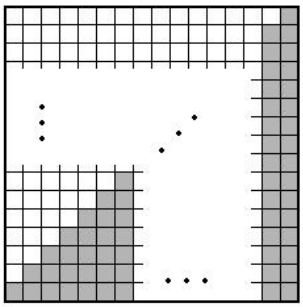
$$(n+1) \cdot n = 2S(n) \quad \text{and therefore,} \quad S(n) = \frac{1}{2}(n+1) \cdot n = \frac{n^2+n}{2}$$

This is the formula we have obtained already, and we found it here in a way that didn't depend on whether n is odd or even! (Look back at the steps we just performed and verify for yourself that the odd/even property of n is really irrelevant.)

Alternate Solution: Visual Diagram

Before we wrap up this puzzle, we want to mention a geometric approach to the solution. We will relate the sum $S(n)$ to the area of a square and find a way

to draw the individual terms of the sum $(1, 2, 3, \ldots, n-1, n)$ as portions of the square's area. Specifically, let's consider an $n \times n$ square and draw the sum's terms as rectangles with increasing height, each with width one unit. See the picture below:



Now, asking for a formula that yields the sum $S(n)$ is equivalent to asking what *area* is covered by all of the rectangles we have drawn inside this square. Trying to add up the individual areas is just restating the puzzle, so we need to think of a way to relate this area to the total area of the square. To do this, let's think about what is left over; that is, how can we describe the area of the square that is *not* covered by the rectangles? Look at the area strictly above the first 1×1 rectangle: it is also a rectangle, with dimensions $(n-1) \times 1$.

Look at the area above the 2×1 rectangle: it is a $(n-2) \times 1$ rectangle. This pattern continues! Eventually, we have a 1×1 rectangle above the $(n-1) \times 1$ rectangle and then no area above the last $n \times 1$ rectangle. What is the total area of all of those rectangles? Well, it looks a lot like the sum $S(n)$ we are considering, but it is just missing the final term, n. Now, we can add up the areas of all the rectangles by relating them to $S(n)$ and then to the square's area:

$$n^2 = S(n) + (S(n) - n) = 2S(n) - n$$

Therefore,

$$S(n) = \frac{n^2 + n}{2}$$

the same formula we had before!

Lessons From This Puzzle

Sometimes there are several completely reasonable ways to approach a puzzle and obtain a solution. Some of them might come to mind first but take longer to execute, some might be trickier to find but lead to a solution much more easily, or some might just go nowhere at all! It's often hard to tell beforehand what's going to happen with any particular approach, so just start trying to work through the puzzle and see what happens, keeping track of what you've

tried and what happened so that you can reassess that approach later. This is a fact we need to keep in mind as we advance in our mathematical careers. We can't always know exactly what to do right away. We are bound to get stuck sometimes, or try roads that end up being dead ends. This shouldn't be discouraging; it's just the way it is!

As a subpuzzle, try redoing the "paired terms" for the odd n case, but instead of leaving off the last term of the sum, try separating the middle term and pairing the numbers from the outside in. Does this give you the same result? Does it seem any easier/faster/different than the approach we used? Alternatively, what if we had handled the even n case by saying that $n = 2k$ for some number k? What would we do for odd n? Does this notation change the procedure at all? Does it make it any easier to handle? Now, can you think of any completely different methods to attack this puzzle?

1.4.3 Some Other Sums

Summing Odd Numbers: Observing a Pattern

While we're on the topic of evaluating sums of integers, let's look at some related problems. First, we will look at an interesting geometric way of interpreting sums of *odd* integers: let's represent 1 as a 1×1 block, and then each successively larger odd integer as a right-angled corner of 1×1 blocks that fits perfectly around the previous such figure. Why would we do this? Well, since these are all odd numbers, successive terms differ in size by two, and lengthening the sides of our corner pieces by one each time allows us to snugly fit the corners around each other and build successively larger squares!

Does this pattern continue? If we believe that it does, how could we prove such a thing? What does this geometric pattern mean in terms of the numerical sums? This is a good question to answer first, because as pretty as the geometric pattern is, it's difficult to work with and manipulate and, ultimately, *prove* decidedly. In essence, pointing to the first few terms of the pattern and saying, "Look, it works!" does *not* constitute an official, mathematical proof, so we must find a better way of formulating this problem. This is not to downplay the meaningfulness and beauty of the pattern we noticed; it is quite interesting that it works out that way and it did provide us with some valuable insight into what might be going on, mathematically, but at the end of the day, that's all it can do for us.

Summing Odd Numbers: Proving Our Findings

Let's try to write the sums represented by the figures above in numerical terms. The corner pieces are made from 1×1 blocks, and there are two more blocks in each corner than the previous one, so each square figure that we saw is represented by a sum like

$$1 \quad \text{or} \quad 1+3 \quad \text{or} \quad 1+3+5 \quad \text{or} \quad 1+3+5+7$$

and so on. What we notice from these terms is that, indeed, they sum to square numbers:

$$1 = 1^2 \qquad 1+3 = 4 = 2^2 \qquad 1+3+5 = 9 = 3^2 \qquad 1+3+5+7 = 16 = 4^2$$

This is really the pattern that we want to prove; it is equivalent to the geometric pattern we noticed before, but now it is written in terms we can manipulate. Let's think about how can we do this, now. Is this pattern similar to anything we've seen before? Have we proved any results about sums of integers? Of course! Look back at the previous puzzle; we proved (in a few ways, in fact) that

$$1+2+3+\cdots+(n-1)+n = \frac{n^2+n}{2}$$

How might this be useful in this puzzle? The sum formula we proved involves *all* consecutive integers from 1 to n, but for the current desired formula, we only want to consider consecutive *odd* integers.

Before, we used the function $S(n)$ to represent the sum of the first n natural numbers, so let's define a function $T(n)$ to represent the sum of the first n odd natural numbers. Now, we need to identify the terms of that sum first, and then relate those to $S(n)$ somehow. Below, we have written out the sums for $n = 1, 2, 3$ and 4. Can you find a way to identify the largest term in the sum and express that in terms of n?

$$n=1: \quad 1, \quad n=2: \quad 1+3, \quad n=3: \quad 1+3+5, \quad n=4: \quad 1+3+5+7$$

Notice that the last term of the sum is always $2n-1$. This is related to a general fact, that any *even* integer can be represented as $2k$, for some particular integer k, and any *odd* integer can be represented as $2n-1$, for some particular integer n. (We can also express an odd integer as $2n+1$ for some integer, as well, right? In this context, it is more convenient to use the $2n-1$ form, though.) Accordingly, we want to find a formula for the sum of the first n odd natural numbers, given by

$$T(n) = 1+3+5+7+\cdots+(2n-3)+(2n-1)$$

Can we relate this sum to $S(n)$ or something similar? Well, notice that the sum

$$S(2n) = 1+2+3+\cdots+(2n-3)+(2n-2)+(2n-1)+2n$$

contains *all* of the natural numbers from 1 to $2n$, whereas $T(n)$ only contains the odd natural numbers in that range. Perhaps it would make sense to subtract those two sums and try to find an expression for the sum of the leftover terms:

$$\begin{aligned} S(2n) - T(n) &= \big(1 + 2 + 3 + \cdots + (2n-1) + 2n\big) \\ &\quad - \big(1 + 3 + 5 + \cdots + (2n-3) + (2n-1)\big) \\ &= 2 + 4 + 6 + \cdots + (2n-2) + 2n \end{aligned}$$

These terms are all of the *even* natural numbers from 2 to $2n$. How can we find a formula for this sum? Do we need to do any extra work, or can we apply a previously-proven result? Well, since all of the terms are *even*, we can divide everything by 2 and write

$$\begin{aligned} \frac{1}{2}\left(S(2n) - T(n)\right) &= \frac{1}{2}\big(2 + 4 + 6 + \cdots + (2n-2) + 2n\big) \\ &= 1 + 2 + 3 + \cdots + (n-1) + n = S(n) \end{aligned}$$

and we can be assured that all of the terms in the sum on the far right are, indeed, integers. Not only that, they are *all* of the consecutive integers from 1 to n, and we have a formula for that sum! Now, *everything is written in terms of formulas we already know*, namely $S(n)$ and $S(2n)$, and the one formula that we are seeking, namely $T(n)$. The last step now is to rearrange the equation to isolate $T(n)$ and then substitute what we know about the formulas involving S:

$$\begin{aligned} \frac{1}{2}\left(S(2n) - T(n)\right) &= S(n) \\ S(2n) - T(n) &= 2S(n) \\ S(2n) - 2S(n) &= T(n) \\ \frac{(2n)^2 + 2n}{2} - \frac{2 \cdot (n^2 + n)}{2} &= T(n) \\ \frac{4n^2 + 2n - 2n^2 - 2n}{2} &= T(n) \\ \frac{2n^2}{2} &= T(n) \\ n^2 &= T(n) \end{aligned}$$

This looks rather nice, doesn't it? Despite having to muddle through some algebraic steps, we arrived at one of the conclusions we were hoping to prove: that the sum of consecutive odd integers is a perfect square. Not only that, we have managed to prove precisely how that square number is related to the number of terms in the sum. Specifically, a concise way of summarizing the result that we just proved is to say that "the sum of the first n odd integers is n^2."

Alternate Solution: An Inductive Argument

Could we have proven this in a different way? What if we had not yet proven the result from the previous section, or if we hadn't thought to use it in that way? Could we have somehow taken advantage of the geometric structure of the sums that we noticed at first?

Let's go back and think about this in a slightly different way. Specifically, let's see why adding one more term to a sum produces another square number. Suppose we knew already that one of the sums produced a square number; we know this is true for the first sum ($1 = 1^2$), but let's assume that this happens for some arbitrary number of terms, n. That is, let's *assume* that

$$1 + 3 + 5 + \cdots + (2n - 3) + (2n - 1) = n^2$$

for some value of n. Given this as a fact, what can we subsequently deduce about the next sum? When we add one more term to the sum, we add on the next odd integer, $2n + 1$, so let's see how this affects the value of the sum:

$$1 + 3 + 5 + \cdots + (2n - 3) + (2n - 1) + (2n + 1) = n^2 + 2n + 1 = (n + 1)^2$$

This seems to confirm our belief, doesn't it? Knowing that one sum behaves in the way we expect it to ("if the sum of the first n odd integers is $n^2 \ldots$") allows us to deduce that the next sum must *also* behave in the same way ("... then the sum of the first $n + 1$ odd integers is $(n + 1)^2$"). Does this also prove the result? What do you think? Does it feel strange to essentially assume our result to prove something further about it? Is that really what we did?

This proof strategy, using one form of the result to prove something about a "subsequent" form of the result is called **mathematical induction**. (In general, the meaning of the term "subsequent" depends on the context; here, it meanis the next sum with one more term.) We will examine this strategy in more detail in the next chapter. For now, we will point out that this is a perfectly sound strategy, but it is highly dependent on the fact that the *first* sum behaves appropriately: $1 = 1^2$. That way, the work we did allows us to deduce that the second sum behaves that way ($1 + 3 = 2^2$), which then alllows us to deduce that the third sum behaves that way ($1+3+5 = 3^2$), and so on... What if we had only been able to prove that second part, but the first sum didn't work out in the way that we wanted? Would we stil be able to prove the result? What does this tell you about the induction strategy, in general? We will address some of these issues in more generality later on.

Generalizing: Arithmetic Series

One final sum problem we want to mention is strongly related to the two that we've seen so far and, in fact, if we had proven this next result first, we wouldn't have had to do anything more to prove the first two! In that sense, this next result is *stronger* than the first two: the truth of this result *implies* the truth of the first two. (This is a common notion in mathematical terminology, to label results as *stronger* or *weaker* than others.)

For this result, we want to examine a general **arithmetic series**. This phrase means that we're adding a sequence of numbers where the difference in value between successive terms is a fixed value. Another way of thinking about this is that each term is obtained from the previous one by adding on a fixed constant. Notice that the sums we've examined in the last two puzzles were arithmetic series: in the first sum, each term differed by 1 (or, we added 1 to each term to get the next term), and in the second sum, each term differed by 2 (or, we added 2 to each term to get the next term).

How can we represent a general arithmetic series? Knowing that successive terms must differ by a fixed constant, let's assign that value a variable, say c, for constant. Now, there must be a first term in the sum, as well, so let's assign that value a variable, say a, since it's the first letter. We just need one more variable to tell us how *many* terms there are in the sum, so we will use k, since we've used that variable before with the same meaning. Now, we can represent the entire sum with just these three variables:

$$A(a,c,k) = a + (a+c) + (a+2c) + (a+3c) + \cdots + (a+(k-2)c) + (a+(k-1)c)$$

We can use the fact that each pair of successive terms differ by c to express the second term using the first term, a, and we can use this to express the third term, and so on, by continually adding c. We wanted k terms in total so, thinking of the first term as $a+0\cdot k$, the final term will be what we obtain after adding c to the first term $k-1$ times (there are k numbers from 0 to $k-1$, inclusive). Notice, as well, that we introduced the notation $A(a,c,k)$ to mean "the sum of the arithmetic series with first term a, constant difference c, and k terms". Now, how can we figure out this sum?

Let's employ a strategy that worked before: in the first sum puzzle, we wrote the terms of the sum forwards and backwards and added them together. This allowed us to create many pairs of terms that all had the same sum, reducing the sum to a multiplication. What happens when we do that here? We see that

$$\begin{array}{ccccccccc}
a & + & (a+c) & + \cdots + & (a+(k-1)c) & = & A(a,c,k) \\
(a+(k-1)c) & + & (a+(k-2)c) & + \cdots + & a & = & A(a,c,k) \\
\hline
(2a+(k-1)c) & + & (2a+(k-1)c) & + \cdots + & (2a+(k-1)c) & = & 2A(a,c,k)
\end{array}$$

Again, we find that each pair of terms has the same sum, and in this case that sum is $2a+(k-1)c$. How many such pairs are there? There are exactly k terms, of course! (That's why we chose to use that variable, even.) Representing the sum as a a multiplication, we can now deduce that

$$2A(a,c,k) = k\cdot(2a+(k-1)c)$$

and therefore,

$$A(a,c,k) = \frac{k}{2}\cdot(2a+(k-1)c)$$

Does this look like what you expected for a result? Did you have any expectations? It sometimes helps to try to "guess" what might happen, and then see if and how the results match up with your intuitions.

Applying the General to the Specific

We mentioned before that the sums we examined previously were both arithmetic series, so does this formula yield the correct value for those sums? In the first puzzle, the values of the variables were $a = 1$, $c = 1$, and $k = n$; plugging in those values yields

$$A(1, 1, n) = \frac{n}{2} \cdot (2 + (n - 1)) = \frac{n}{2} \cdot (n + 1) = \frac{n^2 + n}{2}$$

which is, indeed, what we derived. What about the second sum? What were the values of the variables? Is the formula correct? We will leave it to you to verify that result.

Another Representation

As a final comment for this puzzle, we want to discuss one other way of representing the formula we just derived. Look at the term in the parentheses and write it slightly differently: $a + (a + (k - 1)c)$. Do those terms look particularly interesting? Well, they are the first and last terms of the sum, respectively. This gives us a different way of stating the sum formula we derived: $A(a, c, k) = \frac{k}{2}(a + b)$, where a is the first term of the sum and b is the final term. This version of the formula can be more convenient, and lets us verify some sums more quickly.

For instance, if we told you to find the sum of an arithmetic series with first term 12 and last term 110 and 14 terms in total, you wouldn't have to bother figuring out the constand difference c; instead, you could simply find the sum: $\frac{14}{2} \cdot (12 + 110) = 854$. Much faster, right? What is the value of c for that arithmetic series? Is there an easy way to find c, given a and b and k?

Lessons From This Puzzle

It can be helpful to be aware of previous results, since they can make other proofs shorter and easier. Sometimes, it's difficult to recognize when a particular result would be useful or, even if you recognize its use, it may be difficult to figure out how to apply it. In this case, we recognized that we had proven a sum formula before, so it made sense to at least try to figure out how it might be useful in proving a different sum formula. However, there was a completely different way to prove the formula for the sums of odd integers that didn't depend on our previous result. That hints at a more general result, and a curious mathematician would try to explore the problem in more generality, which we did by looking at an arbitrary arithmetic series. In the end, though, we used multiple strategies for the first two sum formulas, and applied just one of those to the general series problem. Could we have used the strategies in other settings? Could we prove the first sum formula by induction, as well? Could we prove the second sum formula using the forwards/backwards writing technique? Try to use those strategies and see what happens. It may seem strange or unnecessary to you, because we already have the results, but seeing how different techniques work in different settings is a valuable lesson. In mathematics, it is often as

hard (or harder, even) to figure out which strategy to use in a proof as it is to figure out the result to be proven. With that in mind, it's helpful to practice particular strategies to develop an intuition for when they will work and when we need to try something else.

1.4.4 Friend Trends

Problem Statement

This puzzle is based on the following anecdote concerning a Hungarian sociologist and his observations of circles of friends among children.

> "In the 1950s, a Hungarian sociologist S. Szalai studied friendship relationships between children. He observed that in any group of around 20 children, he was able to find four children who were mutual friends, or four children such that no two of them were friends. Before drawing any sociological conclusions, Szalai consulted three eminent mathematicians in Hungary at that time: Erdos, Turan and Sos. A brief discussion revealed that indeed this is a mathematical phenomenon rather than a sociological one. For any symmetric relation R on at least 18 elements, there is a subset S of 4 elements such that R contains either all pairs in S or none of them. This fact is a special case of Ramsey's theorem proved in 1930, the foundation of Ramsey theory which developed later into a rich area of combinatorics."
>
> (Quoted from lecture notes by MIT Prof. Jacob Fox.)

The puzzle we now present follows the same idea but with some smaller numbers. Specifically, we are interested in investigating the smallest size of a group of people that *necessitates* a subgroup of three people that are all mutually friends or all mutually enemies.

> Assume that amongst a group of people, any two of them are either friends or enemies, and that these are the only possible relationships (i.e. no acquaintances or frenemies or anything like that). Take a group of four people and try to assign a designation of friend/enemy to each pair so that there are *no* groups of three people that are all friends or all enemies. Can you do this with a group of five people? How about six? Seven? Ten? Twenty? Try to identify a cutoff number for the size of the group where you can be *guaranteed* to find a subgroup of three people that are all friends or all enemies.

Think carefully about this before turning the page and reading our solution.

Representing The Problem Effectively

Did you figure it out? This is a very tricky puzzle, so don't feel bad if you struggled with finding a solution. In fact, we think that investigating this puzzle is just as important as actually finding an answer, because there are several ways to approach this puzzle and it's always interesting to see how different people interpret the puzzle.

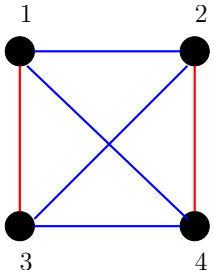
Let's start by discussing how to even write down/draw/talk about this situation. For any of the questions posed by this puzzle, we are meant to consider a group of people with a certain size and think about the relationship between any two people in the group. To tackle this puzzle, we will need a way to represent all of these relationships in an efficient and easily-interpretable way, so that we can verify the desired property about the subgroups of size three. Specifically, we want to easily identify whether or not there are any subgroups of size three that are *homogeneous*, in the sense that the three people are *all* friends or *all* enemies. From now on, we will refer to this as the **"homegeneity property"**.

How could we do this? How could we represent the people and their relatinships? We could number the people in the group, then write out a list of all pairs of numbers and label each pair with F (friend) or E (enemy). Let's try doing that for a group of four people:

$$12F \quad 13E \quad 14F \quad 23F \quad 24E \quad 34F$$

Does this friend/enemy group satisfy the homogeneity property? It's not so easy to verify, is it? For one, the numbering makes it difficult to find subgroups of size three, and to verify the property, we need to check *all* such subgroups to make sure they are not EEE or FFF. Perhaps we should find a better way of *representing* the information of the puzzle before attempting a solution. Can you think of a more visually pleasing way of representing whether two people are friends or enemies, for all possible pairs in the group? Specifically, we would like to have a relatively efficient way of looking for subgroups of size three and recognizing whether they are all friends or all enemies.

Let's try representing each person in the group as a single dot and connecting two people with a type of line dependent on whether those two are friends or enemies. For example, let's connect two people that are friends with a blue line and two people that are enemies with a red line (and remember that any two people are either friends or enemies and nothing else, so each pair of dots must have some colored line between them). For example, the following diagram depicts the relationships we assigned in the line above, with the other notation:



Now, what would we be looking for to verify the homogeneity property? We want three dots (three people) so that all of the lines between them are either blue (all friends) or red (all enemies). That's right—we're looking for **monochromatic triangles**! (Note: we want the vertices of the triangle to be one of the original dots we drew; that is, we don't want a vertex to be a place where two lines cross. Also, *monochromatic* comes from the Greek words *monos* and *khroma*, meaning "one" and "color", respectively.) This representation is much easier to interpret visually and makes checking for a solution much faster.

Based on the diagram above, we have addressed the question regarding four people: we have found a particular arrangement of friends and enemies so that there are no subgroups of size three that are either all friends or all enemies. That is, there are no subgroups of size three with the homogeneity property. This shows that such a situation can be achieved with four people, so we are not *guaranteed* to have a group with the homogeneity property amongst four people.

Can you find another such arrangement? How can you be sure that it's a *different* arrangement than the one we've already seen? How many different arrangements are there that satisfy the homogeneity property? Now, try drawing an arrangment that *does* have a subgroup of size three with the homogeneity property. What does that look like? How many of these arrangements are there?

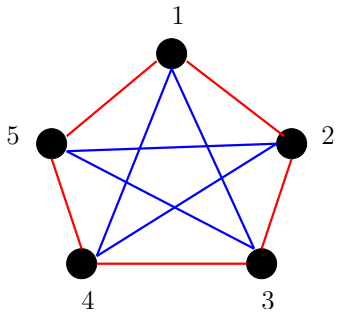
Restating the Problem for $n = 5$

Let's move on and think about a group of five people. Our diagram changes because we have five dots now, and this means there are more lines to draw. Still, we are looking to fill in all of the connections with a blue or red line and make sure there are no monochromatic triangles. Is this possible? (Hint: try arranging the dots into a regular pentagon shape and then filling in the lines.) Try to draw this a few times and see if any of your arrangements work. It may also help to draw in a few lines randomly and then guide your choices from there on out by making sure that you don't create any triangles when you add a new line.

Did you figure it out? Turn the page to see how we did it ...

Solution for $n = 5$

Here is our arrangment of red/blue lines amongst five dots that completely avoids the homogeneity property:



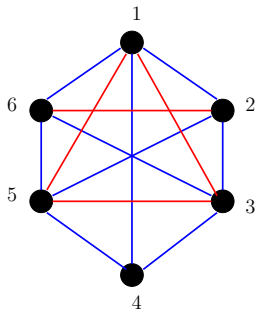
Notice the elegant symmetry of the figure: all of the red lines are on the outside of the pentagon, and all of the blue lines are in the interior of the shape. The reason this works is that any triangle using three dots as vertices must use either two outside lines and one inside line, or two inside lines and one outside line. (Think about that: why couldn't we use three inside lines or three outside lines to make a triangle?) This *guarantees* that any triangle we draw will use two differently-colored lines, so this figure does *not* have the homogeneity property! Of course, we could look at all possible triangles inside the diagram and make sure none of them use one color. How many such triangles are there? How quickly could you check all of them by hand? Is it easier to do that, or to notice the inside/outside property that we mentioned above?

Perhaps you found a solution that doesn't look like the drawing we have. How can you tell whether it's actually a different figure? How many blue and red lines are there in your diagram? In ours? Try redrawing your figure by moving the dots around but maintaining the relationships between the dots (i.e. the color of line drawn between any two of them). Can you make your figure look like ours? What do you think this says about the number of solutions to this puzzle?

What about $n = 6$?

Okay, now we're ready to think about what happens when we have six people. In terms of the dots and lines, we're looking to draw all possible connections between six dots with either blue or red lines and ensure that there are no triangles with the same line type. Before you start drawing, try to think about the solutions to this puzzle when we were working with four and five dots. What did those solutions look like? What was the number of lines that we had to fill in? How many will we have to draw this time? Can we try to make this figure look like the solution for five dots? It sometimes helps to think about how a solution to a current puzzle might be similar to previous work. Now, try to draw this figure and see what happens.

Did it work? Why not? Where did you run into trouble? How many lines could you draw before you were *forced* to make a monochromatic triangle? That is, how many lines could you fit into the figure before the next one you drew would make a monochromatic triangle, no matter whether it was blue or red? These are just tangential questions, in a way, to solving this particular puzzle, but they're worth thinking about because they're interesting in their own right and they may guide us toward a solution for this puzzle or a generalization thereof. For illustration's sake, here is one of our attempts at trying to assing red and blue lines in a diagram. Why did we stop here? How many more lines need to be added? Can we add any of them?

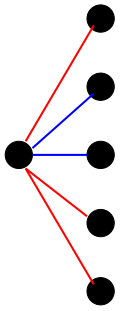


The situation we face now is interesting because it's of the opposite nature to the kind of situations we faced before. With four and five dots, we wanted to show that it was *possible* to arrange all of the lines to make no monochromatic triangles. To show that, we just had to do it! Exhibiting a *particular* figure with the desired property was sufficient to show that it was possible to achieve the property we wanted. With six dots, though, it seems like it is *not* possible to arrange the lines so that there are no monochromatic triangles. How can we prove that this is a fact? It is tempting to say that we should just look at all possible arrangements of the lines and argue that there is *at least* one monochromatic triangle in every single one of them. Is this feasible? How many arrangements of the lines are there? How could we easily find a monochromatic triangle in any given diagram? Remember how we did this with the figure with five dots? We noticed that any triangle would have to use at least one line from the outside and at least one line from the inside, which guarantees right away that any triangle has two types of lines. Could we do the same thing here, and identify some property that *guarantees* a triangle?

The issue is that there are *too many* possible arrangments of the lines in the diagram with six dots for us to check all of them by hand! There are 15 lines to be drawn, each of which could be either blue or red, so it seems like there are 2^{15} possible arrangements. This is a big number! (In actuality, there are slightly fewer possibilities because some of them are equivalent in some sense; more technically, they are called "*isomorphic*".)

Solution: Working with an *Arbitrary* Diagram

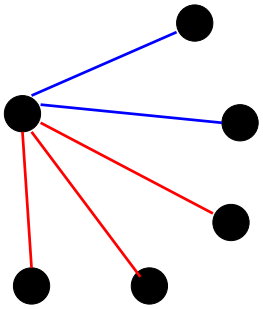
We need to be more clever with our argument so that we can prove a property about *any* diagram without drawing a particular one. That is, we need to find some fact, some property that is true of all of the possible diagrams with six dots, that will still allow us to deduce that there must be a triangle. One way to approach this is to think about drawing the lines in one small section of the diagram. Specifically, let's take any of the six dots and consider the five lines coming out from that dot. For example, we might have something like this:



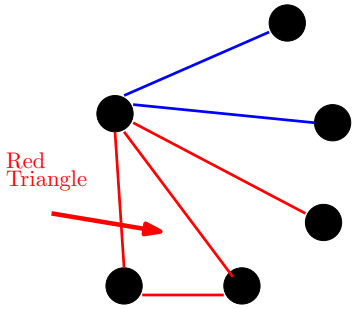
How many are blue and how many are red? This is partially a trick question: we're not really considering any *particular* diagram (like the one above) but rather trying to find a fact about all possible diagrams. Thus, we can't answer that question too specifically. Pretend we are prsented with an **arbitrary** diagram, and we have to make an argument that will work no matter what that diagram was.

Here's what we *can* say: There must be at least three blue lines *or* at least three red lines. Do you see why this is true? The only way this *wouldn't* be true is if there were two (or fewer) blue lines *and* two (or fewer) red lines leaving this particular dot, for a total of four (or fewer) lines. We know that all possible connections must be drawn, though, so there should be five! (This argument is an example of a concept known as the **Pigeonhole Principle**. The idea is that we can't place five objects, of two different colors, into two different boxes without putting three objects of one color into one box. This is an incredibly useful strategy with these types of problems, and we will examine the principle in much greater detail later on in Section 8.6.)

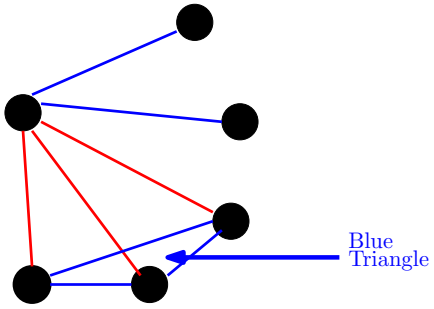
So where do we stand? We started with *any* of the figures with six dots and all lines filled in, and focused on one particular dot; coming out from this dot, there must be three blue lines *or* three red lines. It could be either color, so we can't just assume it's red and follow an argument that way; we can do that, but we have to come back to this point afterwards and see what would change if the three lines were blue. So let's do that: let's examine all of the figures where there are three red lines exiting this particular dot. Where can we go from there? We haven't yet assumed anything about the other lines in the figure, so let's look at what those could be. Examine the picture below to see what line colors we have assumed exist so far:



Now, what lines could be added to this diagram that would avoid making a triangle of one color amongst three dots? We can't necessarily make any assumptions about the lines coming out from the two dots that are isolated in the picture, so let's focus on the three dots on the bottom. What color could the lines among those be? Well, if any of them are red, that would form a monochromatic triangle between the two endpoints of that line and the original dot we focused on! That would be a problem.

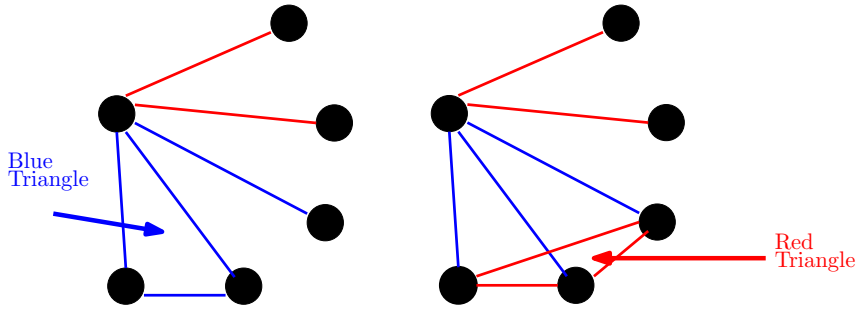


Okay, the only way to avoid that is to make all of those lines blue. But that would make a blue triangle among those three dots! Wow, it looks like we *cannot avoid* making a monochromatic triangle no matter what we do!



Let's go back to our Pigeonhole argument and reassess the situation. What if the three lines of the same type that were guaranteed by the argument were blue instead of red? Well, nothing would really change, right? We would still

be stuck, in terms of adding new lines among the three dots on the bottom of the figure:



If we include any blue lines, that forms a monochromatic triangle with the original dot, and if we make all of them red, that forms a monchromatic triangle there! In this sense, the two arguments we followed after the Pigeonhole argument were *identical*; if we were to replace every instance of the word "blue" with "red" and vice versa, we would have the same argument. Sometimes, mathematicians will use this situation to shorten a proof and just say that "without loss of generality, the three lines are red." This is usually taken to mean that if we were to make the other choice instead (i.e. if the lines were blue) then the further argument would have an identical structure, mathematically, so we can save space and time by not writing the same words over again. This is so common, in fact, that you might sometimes see this phrase, "without loss of generality", abbreviated as **WOLOG** or **WLOG**.

Solution: Summarizing Our Results

What have we accomplished so far? We produced *specific* diagrams that showed we could arrange the lines among four and five dots so that we can avoid a monochromatic triangle, and we argued that *any* diagram of lines with six dots *must* have a monochromatic triangle. In terms of the friends/enemies formulation of the puzzle, this means that any group of six people must have a subgroup of three people that are all friends or all enemies.

Notice how helpful it was to recast the puzzle into this dots/lines formulation; it allowed us to completely forget about the social context of the problem (which can be distracting, in a way) and let us simplify our terminology and notation (we went from labeling pairs of people as "friends" or "enemies" to simply drawing one line between two dots). This is a very useful strategy: extracting the inherent *structure* of a puzzle—the underworkings, the relationships between the elements, how they interact, etc.—and rewriting everything in terms of just those parts. This can make the puzzle easier to understand and tackle, and it can guide us into devising better notation. What if we had continued to solve this puzzle with the $13F, 23E, \ldots$ notation? That may have eventually worked, but it would have been much harder!

One of this puzzle's original questions was to identify a cutoff number so that any larger group of people would necessarily have this subgroup property. Do you think we have accomplished that? Have we identified a cutoff? Could six be that number? Why or why not? In any group of seven people, there's a smaller group of six people, and then our work above proves that there must be three mutual friends/enemies among that group! Certainly, this works for any group of people of size larger than six, so this must be that precise cutoff point we were looking for. This is an analogous result to the one mentioned in the original statement of the puzzle, where the Hungarian sociologist noticed this phenomenon for subgroups of size four. That problem is much more difficult to solve, so we handled a smaller, simpler case here. Both of these results are related to a larger class of problems known as **Ramsey Theory**. This branch of combinatorics and graph theory works with identifying these kind of "cutoff points" where, as the size of some structure (like a group of people) grows and grows, there is eventually a point where we can be *guaranteed* to find a subgroup with a certain property (three mutual friends/enemies). What was at first thought to be a sociological phenomenon turned out to be a rigorous mathematical fact. How about that!

Generalizing: Questions for You

Let's pose some interesting and related questions before moving on. What if we had been looking for homogeneous groups of a different size, like four or five or twelve? Certainly, we would have to have a larger group of people, overall, to guarantee finding such a subgroup. Can we always do this? That is, given any desired subgroup size, can we identify a cutoff point in the way that we did here? Can you figure out how to prove that such a cutoff point *must* exist, even without finding the particular number? Furthermore, what if we allowed for a third possibility: friend, enemy, unacquainted. Could we answer similar questions about homogeneous groups? These are all questions related to Ramsey Theory and some of them are quite difficult to answer and took mathematicians many years to address. Many of these kinds of simply-stated questions are still open, unsolved problems! Don't be discouraged if you don't make any progress on these questions. We believe that even attempting to answer them and thinking about the issues therein is meaningful and beneficial enough, in itself.

Lessons From This Puzzle

This puzzle brought up several difficulties. First, we had to find a way to interpret the puzzle in a meaningful way so that we could even address the questions it asked, and this involved coming up with appropriate *notation* to represent the elements of the puzzle. This is an important part of mathematical problem-solving, particularly for a puzzle like this that doesn't incorporate the notation and visualization as part of the problem statement.

Second, to identify six as the group cutoff size, we had to somehow prove that

something *is not* possible, but the number of possible configurations to examine was way too large to examine each one individually. This happens frequently, particularly in problems related to computer science and algorithms. To address this, we had to employ a strategy far more clever than mere brute force, and it's not always clear what strategy that should be. Here, we essentially started to try to fill in the lines as if it was going to work out, and then realized that we reached a point that was impossible to fix. Proving that something is possible can amount to just showing an example of that phenomenon (which we did with the groups of size four and five), but proving something is *impossible* can be much trickier and require some context-dependent ingenuity.

Lastly, we saw that it can be interesting to think of questions closely-related to the puzzle at hand that simply tweak one or more of the conditions of the problem. What if we look for larger subgroups? What if we allow for more types of lines? How does this change the results? Exploring the boundaries of puzzles by changing the conditions like this can lead to new mathematical discoveries and techniques, and keeps mathematicians actively searching for new knowledge and ways of sharing that knowledge.

1.4.5 The Full Monty Hall

Problem Statement

This puzzle involves only basic probability and arithmetic, yet it has continually stumped some very smart people over the years. In fact, a debate erupted in 1990 when Marilyn vos Savant published the puzzle and her solution in her column in *Parade* magazine, with many people (mathematicians included) writing her letters to agree or disagree with her (correct, we should say) answer. Let's see what you think!

> Suppose you're on a game show and you're given the choice of three doors. Behind one door is a car; behind the others, goats. The car and the goats were placed randomly behind the doors before the show. The rules of the game show are as follows: After you have chosen a door, the door remains closed for the time being. The game show host, Monty Hall, who knows what is behind the doors, now has to open one of the two remaining doors, and the door he opens must have a goat behind it. If both remaining doors have goats behind them, he chooses one at random. After Monty Hall opens a door with a goat, he will ask you to decide whether you want to stay with your first choice or to switch to the last remaining door. Imagine that you chose Door 1 and the host opens Door 3, which has a goat. He then asks you "Do you want to switch to Door Number 2?" Is it to your advantage to change your choice?

Of course, we are assuming that you would prefer to win a car over a goat, and that you want to maximize the chances of winning that one car. Also, we

should mention that this puzzle takes its name from the host (Monty Hall) of a TV game show called *Let's Make a Deal*.

So what do you think? Imagine yourself standing on the stage in front of a TV audience, when Monty Hall asks you, "Do you want to switch to the other door?" What do you do?

Think carefully about this before turning the page and reading our solution.

Solution: *Always* Switch

We'll state the answer right away because it might shock you: you should definitely switch your choice! Reasoning this out and obtaining this solution is the tricky and confusing part, and establishing the right way to interpret the puzzle is part of what has confused solvers for so long.

Analyzing an Incorrect Argument

Let's start by showing you an *incorrect* "solution" that claims switching is irrelevant. Imagine you and your friend heard this puzzle, and he/she gave you this explanation. How would you respond? Would you agree? Why? If not, how would you tell them they're wrong? What is wrong with their explanation?

> Well, after I pick a door and Monty Hall shows me a goat behind another door, then there are only two doors that are unopened. One of them has a goat and the other one has a car, so there's a 50/50 chance that the car is behind my door and a 50/50 chance that the car is behind the other door. Therefore, it doesn't matter whether I switch or not, so I might as well stick with the choice I already made.

Are you convinced by this? Let's try to figure out what's wrong with this argument. The main question we need to address to solve this puzzle involves figuring out two numbers: the probability of winning the car by sticking with our first choice, and the probability of winning the car by switching to the other door. We need to identify these two values and compare them; only *then* can we definitively address the question of the puzzle.

Now, the argument above seems to address both of these probabilities by saying they are both 50%, but there is a problem with how the arguer interprets the situation. What do you think the chances are of winning the car by staying with the first choice? In essence, this is equivalent to not even having Monty Hall show us another door with a goat behind it. If we are going to stay with our first choice of a door, we might as well not even see a goat behind a different door since that doesn't *affect* the object behind the door we originally chose. Let's restate this idea to reiterate its importance:

> Since there are three doors, the chances of picking the right one first is $\frac{1}{3}$, and seeing a goat behind another door *doesn't change that fact.*

This is the problem with the above argument and, in fact, one of the most common mistakes made in "solving" this puzzle.

The next step is to figure out the probability of winning the car *after* switching and compare this to $\frac{1}{3}$. There are several ways to accomplish this, in fact. One concise way is to reason that switching results in success (winning the car) whenever we first choose a door that happens to have a goat. In those cases, the two unchosen doors conceal a goat and a car, in some order, and the game show host is forced to show us the goat; thus, the car is concealed behind the

remaining door, and switching results in a win. Since we will choose a door with a goat behind it $\frac{2}{3}$ of the time, we conclude that switching results in a win $\frac{2}{3}$ of the time.

Enumerating the Possibilities

These explanations might seem unsatisfactory to you, so let's try to actually enumerate (explicitly count) the possible arrangements of the goats and the car behind the doors and write down what happens if we switch in each case. The first thing to notice is that the numbering of the doors is *irrelevant* since all choices are made randomly; that is, whether the car lies behind the door with "#1" printed on it, or "#2" or "#3", the results will be the same: we still have a $\frac{1}{3}$ chance of identifying that door with the car. Accordingly, we can assume WOLOG (remember that this acronym means "without loss of generality") that the car lies behind Door #1 and the goats stand behind Doors #2 and #3. Of course, this is our imposition on the problem, and we can't say that the game-player knows this (otherwise he/she would pick Door #1 every time!). With this arrangement in mind, let's examine all 3 choices we could make at the beginning, and see what switching or staying would accomplish in each case:

Door #1	Door #2	Door #3
Car	Goat	Goat

Our choice	Host shows	Result of switching	Result of staying
Door #1	Door #2 or Door #3	Goat	Car
Door #2	Door #3	Car	Goat
Door #3	Door #2	Car	Goat

One important observation is that when we initially choose the door with the car, the host could choose either of the remaining doors to show us a goat, and he makes that choice *randomly*. For either choice, though, we lose by switching and win by staying. Still, those situations only occur $\frac{1}{3}$ of the time, i.e. after we chose the door with the car behind it, initially. Since each of the rows of the table above is equally likely, we can conclude that $\frac{2}{3}$ of the time we win by switching, while $\frac{1}{3}$ of the time we win by staying.

Does this puzzle make more sense now? Try posing this puzzle to some of your friends and family and gauge their reactions. How many produced the correct answer? How many could correctly explain it? How many erroneously said "it doesn't matter"? How many had already heard the puzzle before?

Generalizing to Many Doors and Cars

Let's look at a generalization of this game show situation and try to prove whether or not switching is also a good idea there. Specifically, let's say there are n doors and m cars in total, so there are $n - m$ goats. To analyze this, we need to specify that $m \leq n - 2$. Think about why this is necessary:

- If $m = n$ were true, then we always win all the time, whether we switch or not. Hence, there's nothing to prove in this case.
- If $m = n - 1$ were true, then whenever we happen to choose the door with the *only* goat behind it at first, the host is *unable* to show us a door with a goat. Hence, the game is ruined and the question of switching is moot.

Now, with these variables in place, here are the new rules of the game: We choose one of the n doors. The host identifies all *other* doors that conceal a goat and randomly chooses one of those doors and opens it. We then have the option to stick with our original choice or switch to *any* of the other doors, of our choosing. What is the strategy now? Should we switch? Should we stay? Does the answer depend on m and n at all? How?

We will approach this modified puzzle in much the same way as the first approach to the version above. We can't possibly enumerate all situations in this version because m and n are unknown variables. Instead, we need to apply logical reasoning to deduce the probabilities associated with winning when staying and when switching. The first key observation is exactly the same as one we made before: the probability of winning when *staying* is exactly the probability of choosing a door concealing a car at first. When we choose a door with a car behind it first, no matter what other door the host reveals, staying on our current choice results in a win. Furthermore, when we chose a door concealing a goat at first, staying results in a loss. Thus, the only way to win while sticking with our first choice is by choosing one of the m doors with a car behind it out of the n total doors. This probability is precisely $\frac{m}{n}$.

To identify the probability of winning after *switching*, we need to think carefully about the probabilities associated with each step. Notice that since $m \geq 2$ is a possibility, it may be that we chose a door with a car at first, switched our choice, and *still won*. With that in mind, we should examine two different cases, here: (a) what happens when we choose a door with a goat first, and (b) what happens when we choose a door with a car first. Each case will leave a different number of options for the host and, subsequently, a different number of ways for us to switch and win, so we should handle them separately.

(a) Let's say we first chose a door with a goat. There are now $n - m - 1$ doors remaining that conceal goats, and the host randomly picks one of those to open. From our perspective, switching leaves us with $n-2$ options (we can't switch to the opened door or our first choice), m of which are cars. Thus, the probability of winning after switching, in this *particular* case, is $\frac{m}{n-2}$.

Since there were $n - m$ goats originally, this case occurs with a probability of $\frac{n-m}{n}$. Therefore, the contribution of this case to the total chances of winning after switching is

$$\frac{n-m}{n} \cdot \frac{m}{n-2} = \frac{nm - m^2}{n(n-2)}$$

(Think about why we *multiplied* these probabilities together. Why did we need to do that at all? Why didn't we add them together? What will we do

to combine this probability with the probability associated with the next case?)

(b) Next, let's say we first chose a door with a car. There are now $n-m$ doors remaining that conceal goats, and the host randomly picks one of those to open. From our perspective, switching leaves us with $n-2$ options, $m-1$ of which are cars. Thus, the probability of winning after switching, in this *particular* case, is $\frac{m-1}{n-2}$.

Since there were m cars originally, this case occurs with a probability of $\frac{m}{n}$. Therefore, the contribution of this case to the total chances of winning after switching is

$$\frac{m-1}{n-2} \cdot \frac{m}{n} = \frac{m^2-m}{n(n-2)}$$

Since these two cases occur separately (i.e. they both can't occur simultaneously) we should add these probabilities together. This will tell us the totaly probability of winning a car after switching from our original choice to another random door:

$$\begin{aligned}\frac{nm-m^2}{n(n-2)} + \frac{m^2-m}{n(n-2)} &= \frac{nm-m^2+m^2-m}{n(n-2)} \\ &= \frac{nm-m}{n(n-2)} \\ &= \frac{m(n-1)}{n(n-2)} \\ &= \frac{m}{n} \cdot \frac{n-1}{n-2}\end{aligned}$$

Now, there's a reason we chose to write the fraction the way we did here. We want to compare this probability to the chances of winning after staying with our first choice of door, which was $\frac{m}{n}$. We see that the probability of winning after switching is, in fact, a multiple of that other probability, and the factor $\frac{n-1}{n-2} > 1$ because $n-1 > n-2$. Written in inequality form:

$$\frac{m}{n} < \frac{m}{n} \cdot \underbrace{\frac{n-1}{n-2}}_{>1}$$

Therefore, the chances of winning after switching are *strictly better* (i.e. not equal to, always better) than the chances of winning after staying. We should always switch to another random door!

Applying the General to the Specific

The original version of this puzzle is the specific case where $n=3$ and $m=1$, so we can check that our result makes sense. The formulas we derived tell us the chances of winning after switching are $\frac{1}{3}$, as we found previously, and the chances of winning after staying are $\frac{1(3-1)}{3(1)} = \frac{2}{3}$, as we found previously. Neat!

Generalizing: Questions for You

What happens for other values of m and n? Can you make the "always switch" strategy significantly better than the "always stay" strategy? That is, how large of a difference can we get between the probability of winning associated with the two strategies? How small can we make it? Is it possible to make them equal?

Another modified version of this puzzle is based on the host opening more than 1 door, revealing multiple goats. Specifically, suppose there are n total doors, m of which have cars, and that, after your first choice, the host randomly identifies p doors with goats and opens all of those, after which you may choose to switch to any of the remaining $n - p - 1$ doors, or stick with your first pick. What is the best strategy in this game? What sorts of conditions do you need to impose on m, n and p to ensure we can even play the game at all? Should you always switch, or does it depend on p? How large/small can we make the difference in the chances of winning for the "always switch" and "always stay" strategies?

Lessons From This Puzzle

Intuition and quick decisions are sometimes helpful in *guiding* us to a solution, but it is always important to check those snap judgments to make sure they are based on sound, rational arguments. In this puzzle, saying that the chances were "50/50" maybe made sense at first, but after thinking about it more carefully and reassessing the situation, we realized there was a flaw in the argument. Specifically, that flaw had to do with interpreting the puzzle correctly and following the steps of the game show in the appropriate order. It was best to assess the probabilities in the order in which they occur as the game is played, rather than starting from an ending position and looking backwards.

In general, puzzles involving probability are quite tricky and require careful analysis, so it's important to keep that in mind. A larger lesson here, as well, is that oftentimes the most simply stated puzzles are the trickiest to figure out. Never be fooled into thinking that a puzzle will be easy to solve because the statement is short or easy to understand!

For more information about the Monty Hall problem and the psychology involved, check out this link to the following paper: Krauss, Stefan and Wang, X. T. (2003). "The Psychology of the Monty Hall Problem: Discovering Psychological Mechanisms for Solving a Tenacious Brain Teaser", *Journal of Experimental Psychology*: General 132(1).

1.5 It's Wise To Exercise

We'll conclude this chapter with a handful of exercises that incorporate some of the ideas we've discussed so far, or give you a chance to practice your previous knowledge, or just keep you on your mental toes. Attempt as many as you can, and discuss potential solutions with some friends to see what they think. At the

end of the day, though, just think of this as a way to keep your brain muscles limber!

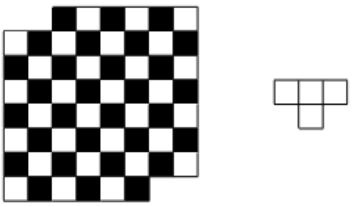
Problem 1.5.1. A fly is resting on the front of a train that is hurtling forward at 60 kilometers per hour. On the same track, 300 kilometers straight ahead, another train is hurtling towards the first train at 60 kilometers per hour. At that moment, when the trains are 300 km apart, the fly takes off at 90 km per hour. He continually flies back and forth between the trains, flying just above the track and instantaneously turning around when he reaches a train. What is the total distance traveled by the fly before the two trains crash together, squishing the fly between them in the process? How did you figure this out? Try to generalize the situation to when one train travels at a km/hr, the other at b km/hr, and the fly at c km/hr.

Problem 1.5.2. A government mint is commissioned to produce gold coins. The mint has 20 machines, each of which is producing coins weighing 5 grams apiece. One day, the foreman of the mint notices that some coins are light, and after assessing the machines, he finds that one of them is making 4 gram coins, while the other 19 machines are working perfectly. He decides to use the situation to his advantage and identify the smartest employee, to be promoted next. He tells the workers that exactly one machine is producing coins that are 4 grams, and that they need to determine which machine is broken. You, as an employee, are allowed to take one, and *only* one, reading on a scale. You can place any number of coins from any machines, of your choosing, but you must pool them together and will only see the total weight of all the coins, as a number in grams. How do you do this so that you can determine precisely which machine is the broken one?

Problem 1.5.3. In a game of chess, the Queen is allowed to move vertically, horizontally, or diagonally, in any direction, for any number of spaces. Try to place 8 Queens on a standard 8×8 chessboard so that *none* of the Queens is attacking any of the others (that is, no Queen can make one move and capture another one immediately). Exhibit a way to do this or show that it is not possible. If you found a way, how many different ways do you think there are to do so? If you showed it isn't possible, find a smaller number of Queens to place so that it is possible. What is the maximum number of Queens you can possibly place on the board in this way?

Problem 1.5.4. Start with a standard 8×8 chessboard and remove 2 squares at opposite corners, say the top-right and bottom-left. Can you cover *all* of the remaining squares with 2×1 dominos so that none of the dominos *overlap*? (Note: this is known as a *tiling* of the board.) Why or why not? What about in the general case, with an $n \times n$ board? Does your answer depend on n at all? How so?

Problem 1.5.5. Consider a standard 8×8 chessboard, with two adjacent squares removed from each corner. (See picture.) Can you tile this board with T-shaped tetrominoes? (See picture.) If so, how? If not, why not?



What about in the general case, with an $n \times n$ board? Does your answer depend on n at all? How so?

Problem 1.5.6. Given a real number x, we let $\lfloor x \rfloor$ denote the largest integer smaller than (or equal to) x and we let $\lceil x \rceil$ denote the smallest integer greater than (or equal to) x. For instance, $\lfloor 6.02 \rfloor = 6$ and $\lfloor 6.99999 \rfloor = 6$ and $\lfloor 6 \rfloor = 6$ and $\lfloor -6.5 \rfloor = -7$.

Determine a more specific and concise representation for the value of the following expressions, whenever possible. (You might have to find a few different expressions, depending on x.)

1. $\lfloor x \rfloor + \lfloor 1 - x \rfloor$
2. $\lceil x \rceil + \lceil 1 - x \rceil$
3. $\lfloor x \rfloor + \lceil x \rceil$
4. $\frac{\lfloor x \rfloor}{x}$
5. $\lfloor x^2 \rfloor - \lfloor x \rfloor^2$
6. $\lceil x^2 \rceil - \lceil x \rceil^2$

Problem 1.5.7. Find three natural numbers a, b, c so that no subset has a sum that is divisible by 3. That is, find a, b, c so that none of the following sums is divisible by 3: $a, b, c, a + b, a + c, b + c, a + b + c$. Is this possible? Why or why not?

Try to do the same thing with 4 numbers: find a, b, c, d so that no subset has a sum that is divisible by 4. Is this possible? Why or why not?

Try to generalize. Can you say anything about finding n natural numbers so that no subset has a sum divisible by n?

Problem 1.5.8. Recall the way that we solved the *Friend Trends* problem, by working with dots and colored lines. For this problem, we want to address a similar situation where we have a certain number of dots and need to draw all possible lines so that any two dots are connected by exactly one line, but in this case we don't care about the color, so we can say all lines are drawn in black, say. Can you draw this figure with 3 dots so that none of the lines cross? What about with 4 dots? Or 5? Or 6? Why or why not? Try to explain why any of those figures are impossible to achieve. If you can't achieve 0 crossings, what is the minimum number you could possibly achieve?

Problem 1.5.9. Let's use the same assumptions from the *Friend Trends* problem: any two people are either friends or enemies, and there are no other relationships. Take a group of n people and assume that no person has more than k friends. We want to separate the n people into some number of clubs so that each person is in exactly one club. How many clubs do we need, in total, so that we can distribute the n people so that nobody is in a club with a friend? Given the relationships between all the people (i.e. given two people, we know whether they are friends or enemies) and that number of clubs, how would we go about separating the people to achieve this property?

Problem 1.5.10. Draw a circle. Place 3 dots along the circumference of the circle. We want to color the sections between the dots (each section gets exactly one color) so that no dot is touched by two sections of the same color. How many colors do we need? What about if we place 4 dots around the circumference of the circle? Or 5? Try to generalize to n dots. What can you say about the number of colors required?

Problem 1.5.11. Suppose you have a drawer full of socks. It contains 2 pairs of blue socks, 3 pairs of red socks, and 4 pairs of green socks. (Also assume that left and right socks are indistinguishable.) One morning, you are in a great rush and start grabbing socks randomly, one at a time, and holding on to all of them until you have a pair. How many socks do you need to take out of the drawer to *guarantee* that you have a pair on your hands?

How does your answer change if we had twice as many pairs, of each color, as before? What if we had 3 pairs of socks in each of red, green, blue, yellow and brown? What if we had n pairs of socks in each of m colors?

Problem 1.5.12. A group of four friends find themselves on one side of a bridge late at night while walking home. The bridge is old and rickety and unsafe to cross as a group. They only have one flashlight between them and the light is only strong enough to light the way for two friends at a time. Each friend has a different level of comfort with the bridge, so they will cross at different speeds. One friend will take 5 minutes to cross, one will take 10, one will take 15, and one will take 20. If two friends cross together, they cross at the pace of the *slower* friend. How long will it take all 4 friends to get across the bridge? Can you find the method that produces the absolute shortest time?

Problem 1.5.13. Consider the usual denominations of coins for the American dollar: pennies (1 cent), nickels (5 cents), dimes (10 cents), and quarters (25 cents). What coins would you need to carry around in your pocket to *guarantee* that you can pay any price between 0 cents and 100 cents with exact change? Are there several possible sets of coins that would achieve this? What is the smallest total value of coins with this property? Are there several possible sets of coins with this same minimum total value?

Problem 1.5.14. Let a, b, c be real numbers, with $a \neq 0$. What is wrong with the following "spoof" that $-\frac{b}{2a}$ is a solution to the equation $ax^2 + bx + c = 0$, if anything?

"Spoof": Let x and y be solutions to the equation. Subtracting $ay^2+by+c = 0$

from $ax^2+bx+c=0$ yields $a(x+y)(x-y)+b(x-y)=0$. Hence, $a(x+y)+b=0$, and so $x+y=-\frac{b}{a}$. Since x and y were *any* solutions, we may repeat this computation with $x=y$. Thus, $2x=-\frac{b}{a}$, and therefore $x=-\frac{b}{2a}$ is a solution. "□"

Problem 1.5.15. Explain why $(-1)(-1)=1$. Pretend you are writing for a fellow classmate of similar intelligence who is skeptical of this fact and needs to be convinced. It is *not enough* to just say "Because it is!" Try to come up with a helpful *geometric* or *physical* explanation, some kind of memorable argument.

Problem 1.5.16. For each of the following proposed equations, identify all of the real numbers that satisfy them:

(1) $|x-2|=|x-3|$

(2) $|2x-1|=|2x-3|$

(3) $|2x-2|=|3x-3|$

(4) $|x+1|=|x-5|$

(5) $|x-1|+|x-2|=|x-3|$

Problem 1.5.17. **The First Rule Of Logic Club Is ...:** To join Logic Club, you must decide to *always* tell the truth or *always* lie. Members of Logic Club know who lies and who soothsays. I do not belong to Logic Club, but I encounter three members on the street who make the following statements:

- Jack: "All three of us are liars."
- Tyler: "Exactly two of us are liars."
- Chuck: "Jack and Tyler are liars."

Who should I believe, if anyone?

Problem 1.5.18. Solve $\sqrt{x-1}=x-3$ for all real values of x that satisfy the equation. Explain your work, and try to indicate why your answer/s is/are the *only* answer/s.

Problem 1.5.19. You have two strings of fuse. Each will burn for exactly one hour. The fuses are not necessarily identical, though, and do not burn at a constant rate. All you have with you is a lighter and these two fuses. Can you measure exactly 45 minutes? If so, explain how. If not, explain why.

Problem 1.5.20. This problem is a variation of a standard puzzle, a form of which first appeared in the Saturday Evening Post way back in 1926!

Three friends pitched in and bought a big bag of M&M candies. They brought the box back to their apartment and decided to wait and share the bag the next day at a party.

During the night, the first friend woke up and felt like having a snack. He

decided to just eat his share of the candies now and not have any the next day. He opened the bag, divided the M&Ms into three equal piles, but realized there was one left over. He figured that one extra couldn't hurt, and ate his share and the extra, then put the rest back in the bag.

Later in the night, the second friend did the exact same thing. He woke up feeling hungry, divded what was left in the bag into three piles, and ate his share plus the extra one that was left over.

Even later in the night, the third friend did the exact same thing, including the extra M&M left over.

The next day at their party, the friends split the bag of remaining candies into three *equal* shares and enjoyed them. (No one acknowledged what they had all done, of course).

How many M&Ms were in the bag to begin with? What is the smallest possible number?

Problem 1.5.21. Given a list of real numbers, their *arithmetic mean* is defined to be their sum divided by the number of terms, and their *geometric mean* is defined to be their product raised to the power of one over the number of terms. That is, supposing we have $x_1, x_2, \ldots, x_n$ that are real numbers, then the arithmetic mean is

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

and the geometric mean is

$$\sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

(Note: The n-th root of a number is the same as that number raised to the power of $\frac{1}{n}$.)

Can you identify two numbers so that the arithemtic and geometric means are *equal*? Can you identify two numbers so that the arithmetic mean is strictly greater than the geometric mean? How about the other way around?

Repeat this with three numbers, four numbers, etc. Can you identify a general pattern?

Problem 1.5.22. Consider the variable equation $6x + 15y = 93$. We want to find some *integral* solutions; that is, we want to find values of x and y that are both *integers* (natural numbers, zero, and negative natural numbers) that satisfy the equation.

1. Find a solution where both x and y are positive integers. Decribe, with a few sentences, how you came up with this solution.
2. Find a solution where one of the values, x or y, is positive and the other is negative. Again, describe how you came up with this solution.
3. How many solutions do you think there are? Try to write down a characterization of all possible solutions, or describe how you might find them all.

Problem 1.5.23. A **magic square** is an $n \times n$ array that contains each of the numbers from 1 to n^2 exactly once and has the property that every row and column (and both of the main diagonals) sums to the same number.

For example, here is a 3×3 magic square:

8	1	6
3	5	7
4	9	2

Notice that the so-called **magic sum** of each row/column/diagonal is 15 in this case.

Can you find a formula for what the magic sum of an $n \times n$ magic square must be?

(*Hint:* There is a result we discovered in this chapter that will be useful.)

Problem 1.5.24. How many whole numbers less than or equal to 1000 have a 1 as at least one of their digits? For instance, we want to count 1 and 12 and 511 once each.

Problem 1.5.25. We have several piles of koala bears. In an attempt to disperse them, we remove exactly one koala bear from each pile and place all of those koalas into one new pile. For example, if we started with koala piles of sizes 1, 4, and 4, we would then end up with koala piles of sizes 3, 3, and 3; or, if we started with piles of size 3 and 4, we would end up with piles of size 2 and 3 and 2.

It **is** possible that we do this operation *exactly once* and end up with the *exact same pile sizes* as we started with (the order of them is irrelevant; only the *sizes* matter).

Identify all the collections of piles that have this property and explain why they are the only ones.

Hint: An example of a starting situation with this property is when we have just one pile of size 1. We do the operation and again obtain one pile of size 1. Bingo.

Hint 2: Be sure to also explain why your situations are the *only* ones that work. How can we be sure you didn't miss some answers?

1.6 Lookahead

This introductory chapter is meant to get you thinking about what **mathematics** is, how we **solve problems**, and what it means to write a **proof**. Throughout the rest of the book, we will be discussing all three of these ideas in more and more detail. In doing so, we will explore several different areas of the mathematical universe. We have an overarching plan for our journey, so don't think that we are just stumbling randomly through the forest. Our major

goals are to (a) formalize some of the intuitive ideas we have about mathematical objects, (2) see many examples of good proofs and develop the ability to create and write good proofs, (3) develop problem-solving skills and the ability to apply mathematical knowledge, and (4) cultivate an appreciation for both the art and science of mathematics.

Scan the table of contents at the beginning of the book to get a sense for where our journey is headed. The phrases and terminology might be foreign to you now, but by the end of the book, we will all be speaking the same language: **mathematics**.

Chapter 2

Mathematical Induction: "And so on . . ."

2.1 Introduction

This chapter marks our first big step toward investigating mathematical proofs more throughly and learning to construct our own. It is also an introduction to the first significant **proof technique** we will see. As we describe below, this chapter is meant to be an appetizer, a first taste, of what **mathematical induction** is and how to use it. A couple of chapters from now, we will we be able to rigorously define induction and *prove* that this technique is mathematically valid. That's right, we'll actually prove how and why it works! For now, though, we'll continue our investigation of some interesting mathematical puzzles, with these particular problems hand-picked by us for their use of inductive techniques.

2.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to ...

- Define what an inductive argument is, as well as classify a presented argument as an inductive one or not.
- Decide when to use an inductive argument, depending on the structure of the problem you are solving.
- Heuristically describe mathematical induction via an analogy.
- Identify and describe different kinds of inductive arguments by comparing and contrasting them, as well as identify the underlying structures of the corresponding problems that would yield these similarities and differences.

2.1.2 Segue from previous chapter

As in the previous chapter, we won't assume any familiarity with more advanced mathematics beyond basic algebra and arithmetic, and perhaps some visual, geometric intuition. We will, however, make use of summation and product notation fairly often, so if you feel like your notational skills are, go back and review Section 1.3.5.

2.1.3 Motivation

Look back at the Puzzle in Section 1.4.3, where we proved that the sum of the first n odd natural numbers is exactly n^2. We first observed this pattern geometrically, by arranging the terms of the sums (odd integers) as successively larger "corner pieces" of a square. The first way we proved the result, though, didn't seem to depend on that observation and merely utilized a previous result (about sums of even *and* odd integers) in an *algebraic* way; that is, we did some tricky manipulation of some equations (mutliplying and subtracting and what have you) and then–voilà!–out popped the result we expected. What did you think about that approach? Did it feel satisfying? In a way, it didn't quite match the geometric interpretation we had, at first, so it might be surprising that it worked out so nicely. (Perhaps there is a *different* geometric interpretation of this approach. Can you find one?)

Our second approach was to model that initial geometric observation. We transformed visual pieces into algebraic pieces; specifically, a sum was related to the area of a square, and the terms of the sum were related to particular pieces of that square. We established a *correspondence* between different intepretations of the same problem, finding a way to relate one to the other so that we could work with either interpretation and know that we were proving something about the overall result. The benefit of the visual interpretation is that it allowed us to take advantage of a general proof strategy known as **mathematical induction**, or sometimes just **induction**, for short. (The word *induction* has some non-mathematical meanings, as well, such as in electromagnetism or in philosophical arguments, but within the context of this book, when we say *induction*, we mean

mathematical induction.) What exactly is induction? How does it work? When can we use this strategy? How do we adapt the strategy to a particular puzzle? Are there variations of the strategy that are more useful in certain situations? These are all questions that we hope to answer in this chapter.

The first topic we'd like to address is a question that we didn't just ask in the last paragraph, namely, "*Why* induction? *Why* bother with it?" Based on that puzzle in Section 1.4.3, it would seem that mathematical induction isn't entirely necessary since there might be other ways of proving something, instead of by induction. Depending on the context, this very well may be true, but the point we'd like to make clear from the beginning is that *induction is incredibly useful!* There are many situations where a proof by induction is the most concise and clear approach, and it is a well-known general strategy that can be applied in a variety of such situations. Furthermore, applying induction to a problem requires there to be a certain *structure* to the problem, a dependence of one "part" of the result on a "previous part". (The "parts" and the "dependence" will depend on the context, of course.) Recognizing that induction applies, and actually going through the subsequent proof process, will usually *teach* us something about the inherent structure of the problem. This is true even when induction fails! Perhaps there's a particular part of a problem that "ruins" the induction process, and identifying that particular part can be helpful and insightful.

We hope to motivate these points through some illustrative examples first, after which we will provide a reasonably thorough *definition* of mathematical induction that will show how the method works, in generality. (A completely *rigorous* definition will have to be put off until a little bit later, after we have defined and investigated some relevant concepts, like set theory and logical statements and implications. For now, though, the definition we give will suffice to work on some interesting puzzles and allow us to discuss induction as a general proof strategy.)

2.1.4 Goals and Warnings for the Reader

Do keep in mind that we are still building towards our goal of mathematical rigor, or as much as is possible within the scope and timing of this book and course. Some of the claims we make in this Chapter will be clarified and technically proven later on, once we have properly discussed the natural numbers and some basic mathematical logic. All in due time!

That said, this chapter is still very important, since we are continuing to introduce you to the process of solving mathematical problems, applying our existing knowledge and techniques to discover new facts and explain them to others. In addition, mathematical induction is a fundamental proof technique that will likely appear in every other mathematics course you take! This is because of its usefulness and the prevalence of inductive properties throughout the mathematical world.

2.2 Examples and Discussion

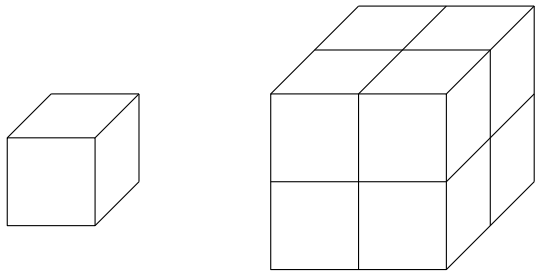
2.2.1 Turning Cubes Into Bigger Cubes

To motivate the overall method of mathematical induction, let's examine a geometric puzzle and solve it together. This example has been chosen carefully to illustrate how **mathematical induction** is relevant when a puzzle has a particular type of structure; specifically, some truth or fact or observation *depends* or *relies* or can be *derived* from a "previous" fact. This dependence on a previous case (or cases) is what makes a process *inductive*, and when we observe this phenomenon, applying *induction* is almost always a good idea.

1-Cube into a 2-Cube

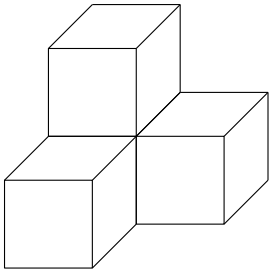
Let's examine cubic numbers and, specifically, let's try to describe a cubic number *in terms of* the previous cubic number. Imagine a $1 \times 1 \times 1$ cube, just one building block. How can we build the "next biggest" cube, of size $2 \times 2 \times 2$, by adding $1 \times 1 \times 1$ building blocks? How many do we need to add? Arithmetically, we know the answer: $2^3 = 8$ and $1^3 = 1$, so we need to add 7 blocks to have the correct volume. Okay, that's a specific answer, but it doesn't quite tell us *how* to arrange those 7 blocks to make a cube, nor does it give us any insight into how to answer this question for *larger* cubes. Ultimately, we would like to say how many blocks are required to build a $100 \times 100 \times 100$ cube into a $101 \times 101 \times 101$ cube without having to perform a lot of tedious arithmetic; that is, we are hoping to eventually find an answer to the question: given an $n \times n \times n$ cube, how many blocks must we add to build it into a $(n+1) \times (n+1) \times (n+1)$ cube? With that in mind, let's think carefully about this initial case and try to answer it with a general argument.

Given that single building block, and knowing we have to add 7 blocks to it, let's try to identify exactly where those 7 blocks should be placed to make a $2 \times 2 \times 2$ cube. (For simplicity, we will refer to a cube of size $n \times n \times n$ as an n-cube, for any value of n. We will only need to use values of n that are *natural numbers*, i.e. non-negative whole numbers, in this example.) Look at the pictures of the 1-cube and 2-cube below and try to come up with an explanation of constructing one from the other.

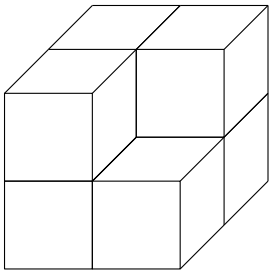


Here's one reasonable explanation that we want to use because it will guide us in the general explanation of building an $(n+1)$-cube from an n-cube, and because

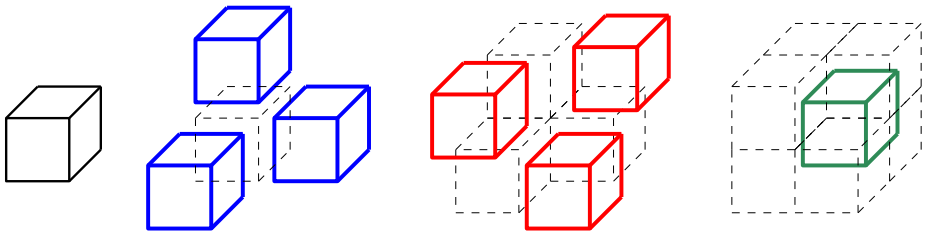
it is a mathematically *elegant* and simple explanation. Start with the 1-cube positioned as it is above and "enlarge" the 3 exposed faces by the appropriate amount, in this case by one block. This accounts for 3 of the 7 blocks, thus far: $2^3 = 1^3 + 3 + ___$. Where are there "holes" now?



The blocks we just added have created "gaps" between each pair of them, and each of those "gaps" can be filled with one block. This accounts for 3 more of the 7 total blocks: $2^3 = 1^3 + 3 + 3 + ___$. Now what?



There is just one block left to be filled, and it's the very top corner. Adding this block completes the 2-cube and tells us how to mathematically describe our construction process with the following picture and equation:

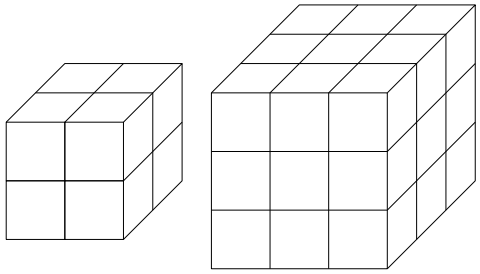


$$2^3 = 1 + 3 + 3 + 1$$

2-Cube into a 3-Cube

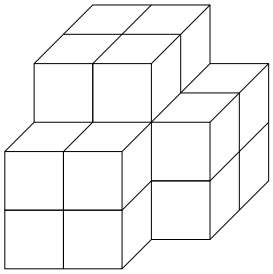
Okay, we might now have a better idea of how to describe this process in general, but let's examine another case or two just to make sure we have the full idea.

Let's start with a 2-cube and construct a 3-cube from it. (You can even try this out by hand if you happen to own various sizes of Rubik's Cubes!) We can follow a process similar to the steps we used in the previous case and just change the numbers appropriately. Starting with a similar picture



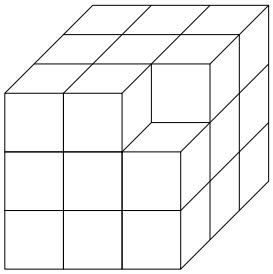
we see that we need to "enlarge" the three exposed faces of the 2-cube but, in this case, the amount by which we need to enlarge them is *different* than before (with the 1-cube) since we are working with a larger initial cube. Specifically, each face must be enlarged by a 2×2 *square* of blocks (whereas, in the previous case, we added a 1×1 square of blocks). Thus, an equation to account for this addition is

$$3^3 = 2^3 + 3 \cdot 2^2 + ___$$

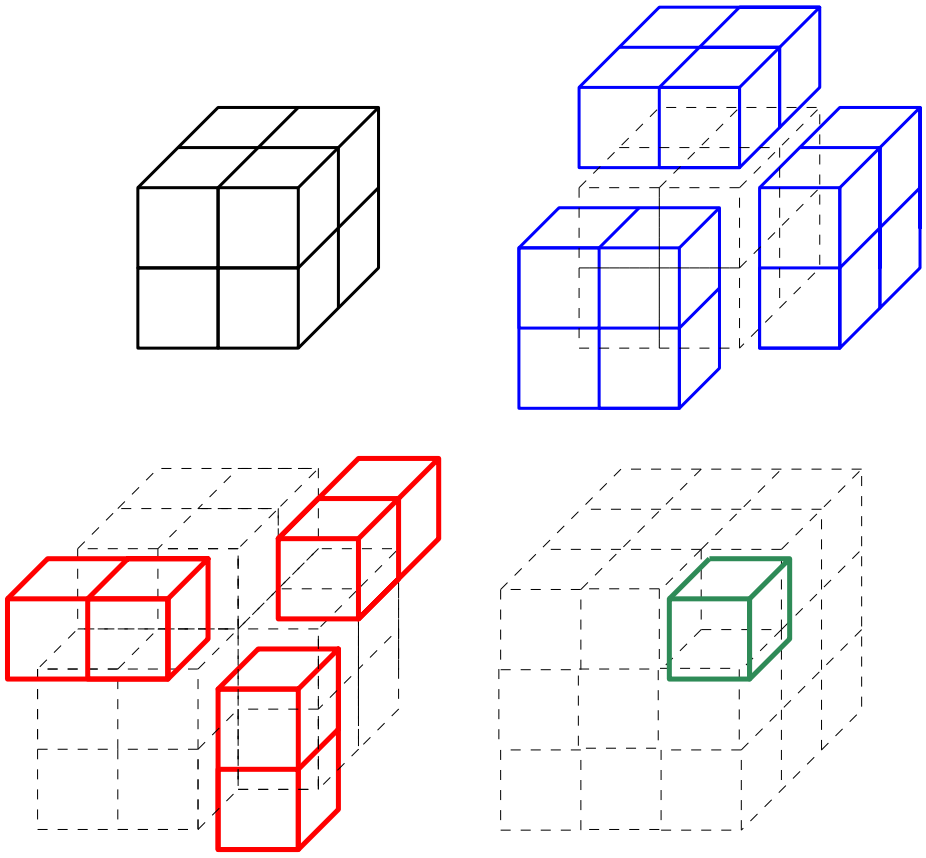


After we do this, we see that we need to fill in the gaps between those enlarged faces with 2×1 of blocks (whereas, in the previous case, we added 1×1 rows of blocks). An equation to account for the additions thus far is

$$3^3 = 2^3 + 3 \cdot 2^2 + 3 \cdot 2 + ___$$



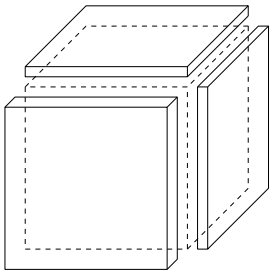
After we do this, we see that there is only the top corner left to fill in. Accordingly, we can depict our construction process and its corresponding equation:



$$3^3 = 2^3 + 3 \cdot 2^2 + 3 \cdot 2 + 1$$

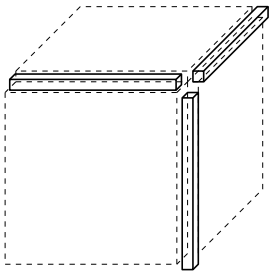
n-Cube into an $(n+1)$-Cube

Do you see now how this process will generalize? What if we started with an n-cube? How could we construct an $(n+1)$-cube? Let's follow the same steps we used in the previous two cases. First, we would enlarge the three exposed faces by adding three *squares* of blocks. How big is each square? Well, we want each square to be the same size as the exposed faces, so they will be $n \times n$ squares, accounting for n^2 blocks for each face:



$$(n+1)^3 = n^3 + 3n^2 + ___$$

Next, we would fill in the gaps between these enlarged faces with rows of blocks. How long are those rows? Well, they each lie along the edges of the squares of blocks we just added, so they will each be of size $n \times 1$, accounting for n blocks for each gap:



$$(n+1)^3 = n^3 + 3n^2 + 3n + ___$$

Finally, there will only be the top corner left to fill in! Therefore,

$$(n+1)^3 = n^3 + 3n^2 + 3n + 1$$

"Wait a minute!" you might say, abruptly. "We already knew that, right?" In a way, yes; the equation above is an algebraic identity that we can also easily see by just expanding the product on the left and collecting terms:

$$\begin{aligned}(n+1)^3 &= (n+1)\cdot(n+1)^2 \\ &= (n+1)\cdot(n^2+2n+1) \\ &= (n^3+2n^2+n)+(n^2+2n+1) \\ &= n^3+3n^2+3n+1\end{aligned}$$

So what have we really accomplished? Well, the main point behind deriving this identity in this geometric and visual way is that it exhibits how this identity represents some kind of *inductive* process. We sought to explain how to derive one "fact" (a cubic number, $(n+1)^3$) from a previously known "fact" (the next smallest cubic number, n^3) and properly explained how to do just that. Compare this to one of the methods we used to investigate the fact that the sums of odd integers yield perfect squares, too. That observation also belies an

inductive process and, although we didn't describe it as such at the time, we encourage you to think about that now. Look back at our discussion and try to write out how you could write $(n+1)^2$ in terms of n^2 by looking at squares of blocks. Does it look anything like an "obvious" algebraic identity? (If you're feeling ambitious, think about what happens with writing $(n+1)^4$ in terms of n^4. Is there any geometric intuition behind this? What about higher powers?)

The benefit of the method we've used is that we now know how to describe cubic numbers in terms of smaller cubic numbers, *all the way down to* 1; that is, any time we see a cubic number in an expression, we know precisely how to write that value in terms of a smaller cubic number and some leftover terms. Furthermore, each of those expressions and leftover terms have an inherent *structure* to them that depends on the cubic number in question. Thus, by iteratively replacing any cubic number, like $(n+1)^3$, with an expression like the one we derived above, and continuing until we can't replace any more, should produce an equation that has some built in *symmetry*. This idea is best illustrated by actually doing it, so let's see what happens. Let's start with the expression we derived, for some arbitrary value of n,

$$(n+1)^3 = n^3 + 3n^2 + 3n + 1$$

and then recognize that we now know a similar expression

$$n^3 = (n-1)^3 + 3(n-1)^2 + 3(n-1) + 1$$

We proved that this equation holds when we gave a general argument for the expression above for n^3, since that only relied on the fact that $n \geq 1$. We can follow the same logical steps, throughout replacing n with $n-1$, and end up with the second expression above, for $(n-1)^3$. (Does this keep going, for *any* value of n? Think about this for a minute. Does our argument make any sense when $n \leq 0$? Would it make physical sense to talk about, say, constructing a $(-2) \times (-2) \times (-2)$ cube from a different cube?)

Therefore, we can replace the n^3 term in the line above

$$\begin{array}{rcccccccc} (n+1)^3 = & & \cancel{n^3} & + & 3n^2 & + & 3n & + & 1 \\ & + & (n-1)^3 & + & 3(n-1)^2 & + & 3(n-1) & + & 1 \end{array}$$

This is also an algebraic identity, but it's certainly not one that we would easily think to write down just by expanding the product on the left-hand side and grouping terms. Here, we are taking advantage of the *structure* of our result to apply it over and over and obtain new expressions that we wouldn't have otherwise thought to write down. Let's continue with this substitution process and see where it takes us! Next, we replace $(n-1)^3$ with the corresponding expression and find

$$\begin{array}{rcccccccc} (n+1)^3 = & & & & 3n^2 & + & 3n & + & 1 \\ & & \cancel{(n-1)^3} & + & 3(n-1)^2 & + & 3(n-1) & + & 1 \\ & + & (n-2)^3 & + & 3(n-2)^2 & + & 3(n-2) & + & 1 \end{array}$$

Perhaps you see where this is going? We can do this subsitution process over and over, and the columns that we've arranged above will continue to grow, showing us that there is something deep and mathematically symmetric going on here. But where does this process stop? We want to write down a concise version of this iterative process and be able to explain all of the terms that arise, so we need to know where it ends. Remember the very first step in our investigation of the cubic numbers? We figured out how to write $2^3 = 1^3 + 3 + 3 + 1$. Since this was our *first* step in *building* this inductive process, it should be the *last* step we apply when building backwards, as we are now. Accordingly, we can write

$$\begin{array}{rcccccccc}
(n+1)^3 = & & & & 3n^2 & + & 3n & + & 1 \\
& & + & & 3(n-1)^2 & + & 3(n-1) & + & 1 \\
& & + & & 3(n-2)^2 & + & 3(n-2) & + & 1 \\
& & + & & 3(n-3)^2 & + & 3(n-3) & + & 1 \\
& & & & \vdots & + & \vdots & + & \vdots \\
& & + & & 3 \cdot 2^2 & + & 3 \cdot 2 & + & 1 \\
+ & 1^3 & + & & 3 \cdot 1^2 & + & 3 \cdot 1 & + & 1
\end{array}$$

This is *definitely* an identity we wouldn't have come up with off the top of our heads! In addition to being relatively pretty-looking on the page like this, it also allows us to apply some of our previous knowledge and simplify the expression. To see how we can do that, let's apply summation notation to the columns above and collect a bunch of terms into some simple expressions:

$$(n+1)^3 = 1^3 + 3 \cdot \sum_{k=1}^{n} k^2 + 3 \cdot \sum_{k=1}^{n} k + \sum_{k=1}^{n} 1$$

In the last chapter, we saw a couple of different proofs that told us

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

Let's use that fact in the line above, and also simplify the term on the far right, to write

$$(n+1)^3 = 1 + 3 \cdot \sum_{k=1}^{n} k^2 + \frac{3n(n+1)}{2} + n$$

What does this tell us? What have we accomplished after all this algebraic manipulation? Well, we previously proved a result about the sum of the first n natural numbers, so a natural question to ask after that is: What is the sum of the first n natural numbers *squared*? How could we begin to answer that? That's a trick question, because *we already have*! Let's do one or two more algebraic steps with the equation above by isolating the summation term and

then dividing:

$$(n+1)^3 - 1 - n - \frac{3n(n+1)}{2} = 3 \cdot \sum_{k=1}^{n} k^2$$

$$\frac{1}{3}(n+1)^3 - \frac{1}{3}(n+1) - \frac{n(n+1)}{2} = \sum_{k=1}^{n} k^2$$

This is what we've accomplished: we've derived a formula for the sum of the first n square natural numbers! Of course, the expression on the left in the line above isn't particularly nice looking and we could perform some further simplification, and we will leave it to you to verify that this yields the expression below:

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

"And so on" is not rigorous!

There are a couple of "morals" that we'd like to point out, based on all of this work. The first moral is that generalizing an argument is a good method for discovering new and interesting mathematical ideas and results. Did you think about how this puzzle is related to the sums of odd natural numbers? If not, we encourage you strongly to try that now, as well as think about generalizing this even further to four or five dimensional "cubes" and so on. In addition to giving you some other interesting results, it will also be incredibly instructive for learning to think abstractly and apply inductive processes. The second moral is more like an admission: we have *not* technically *proven* the formula above for the sum of the first n square natural numbers. It seems like our derivation is valid and tells us the "correct answer" but there is a glaring issue: ellipses! In expanding the equation for $(n+1)^3$ and obtaining those columns of terms that we collected into particular sums, writing $\vdots$ in the middle of those columns was helpful in guiding our intuition, but *this is not a mathematically rigorous technique.* How do we *know* that all of the terms in the middle are exactly what we'd expect them to be? How can we be so sure that all of our pictures of cubes translate perfectly into the mathematical expressions we wrote down? What do we really mean by "and keep going all the way down to 1"?

As an example, consider this:

$$1, 2, 3, 4, \ldots, 100$$

What is that list of numbers? You probably interpreted it as "all the natural numbers between 1 and 100, inclusive". That seems reasonable. But what if we *actually* meant this list?

$$1, 2, 3, 4, 7, 10, 11, 12, 14, \ldots, 100$$

Why, of course, we meant the list of natural numbers from 1 to 100 that don't have an "i" in their English spelling! Wasn't it obvious?

The point is this: when talking with a friend, and *verbalizing* some ideas, it might be okay to write "$1, 2, 3, \ldots, 100$" and ensure that whoever is listening knows *exactly* what you mean. In general, though, we can't assume that a reader would just naturally *intuit* whatever we were trying to convey; we should be as *explicit* and *rigorous* as possible.

It may seem to you now like we're nit-picking, but the larger point is that there is a mathematical way of making this argument more *precise*, so that it constitutes a completely valid *proof*. Everything we have done so far is useful in guiding our intuition, but we will have to do a little more work to be sure our arguments are completely convincing. There are a few other concepts required to make this type of argument rigorous, in general, and we will investigate those in the next chapter and return to this subject immediately after that. However, in the meantime, let's examine one more example to practice this intuitive argument style and recognizing when induction is an applicable technique.

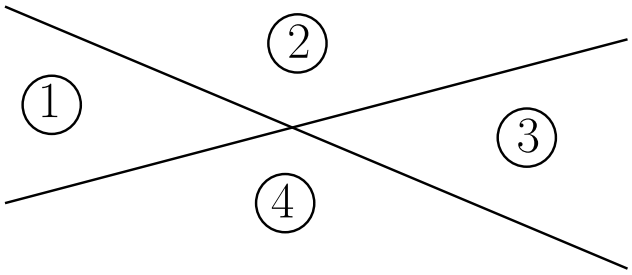
2.2.2 Lines On The Plane

Take a clean sheet of paper and a pen and a ruler. How many regions are on your sheet? Just one, right? Draw a line all the way across the paper. Now there are two regions. Draw another straight line that intersects your first line. How many regions are there? You should count four in total. Draw a third line that intersects both of the first two, but *not* at the point where the first two intersect. (That is, there should be three intersection points, in total.) How many regions are there? Can you predict the answer before counting? What happens when there are 4 lines? Or 5? Or 100? How do we approach this puzzle and, ultimately, solve it? Let's give a more formal statement to be sure we're thinking the same way:

Consider n lines on an infinite plane (two-dimensional surface) such that no two lines are *parallel* and no more than two lines *intersect* at one point. How many distinct regions do the lines create?

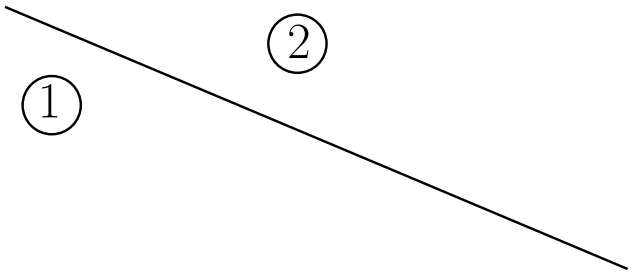
We can draw a few examples by hand when n is small (up to, say, $n = 5$ is reasonable), and let's use this to guide our intuition into making a general argument for an *arbitrary* value of n. (Notice that this strategy is very similar to what we did in the previous puzzle: identify a pattern with small cases, identify the relevant components of those cases that can generalize, then abstract to an aribtrary case.) Specifically, we want to attempt to identify how the number of regions in one drawing *depends* on the number of regions in a drawing with fewer lines. What happens when we draw a new line? Can we determine how this changes the already existing regions? Can we somehow count how many regions this creates? Do some investigation of this puzzle on your own before reading on. If you figure out some results, compare your work to the steps we follow below.

Let's start with a small case, say $n = 2$. We know one line divides a plane into 2 regions; what happens when we add a second line? We know we get 4 regions, because we can just look and count them:

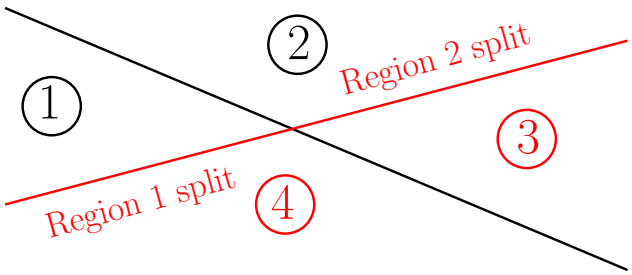


However, we are only looking at *one specific case* of two intersecting lines. How do we know that we will *always* find four regions, no matter how we draw those two lines appropriately? That is, can we describe *how* this happens in a way that somehow incorporates the fact that the number of liens is $n = 2$? Think about it!

Here's our approach. Notice that each of the already existing regions is split into two when we add a second line, and that this is true *no matter how you choose to draw the lines*; as long as we make sure the two lines aren't parallel, they will always behave this way. That is, if we take one line that splits the plane into two regions,

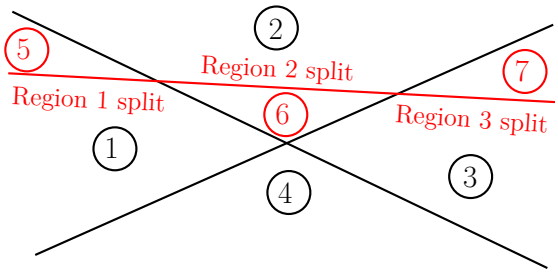


then adding a new line will split each of those existing regions in two. This adds two new regions to the whole plane, giving four regions in total:

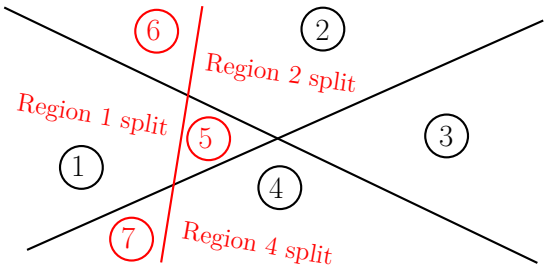


What about when $n = 3$? In this case, we want to think about adding a third line to a diagram with two lines and four regions. We want to make an argument that doesn't depend on a *particular* arrangement of the lines, so eventually the only facts we should use are that no lines are *parallel* and any point of *intersection* only lies on two lines (not three or more). For now, though,

it helps to look at a particular arrangement of lines so that we are talking about the same diagram; we can use our observations of this specific diagram to guide our general argument. Let's start with a two-line diagram, on the left below, and add a third line, but let's choose the third line so that all of the intersection points are "nearby" or within the scope of the diagram, so that we don't have to rescale the picture:

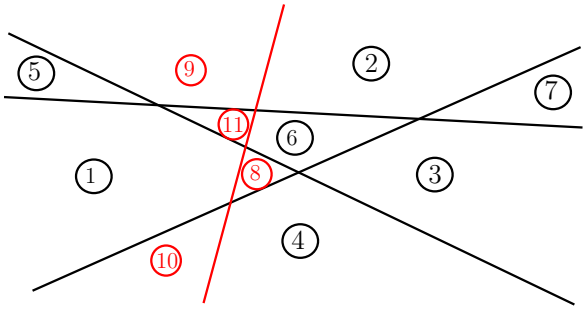


We certainly have 7 regions now, but we made the third line a separate color so that we can identify where the "new" regions appear: one region (the lower quadrant, Region 4) remains unchanged, but the three other regions are split in two and each of those "splits" adds 1 to our count (where there was 1 region, now there are 2). What if we had placed the line differently?



The same phenomenon occurs, where one quadrant remains untouched but the other three are split in two. (How do we know there aren't any other regions not depicted within the scale of our diagram? This is not as easy a question to answer as you might think at first blush, and it's worth thinking about.) Experiment with other arrangements of the three lines and try to convince yourself that this always happens; also, think about *why* this is the case and *how* we could explain that this must happen. Before giving a general explanation, though, let's examine another small case.

When $n = 4$, we start with three lines and 7 regions and add a fourth line that is not parallel to any of the existing lines and doesn't pass through any existing intersection points. Again, we will want to make an argument that isn't tied to a specific arrangement of the lines, but looking at the following specific diagram will help guide our intuition into making that argument:

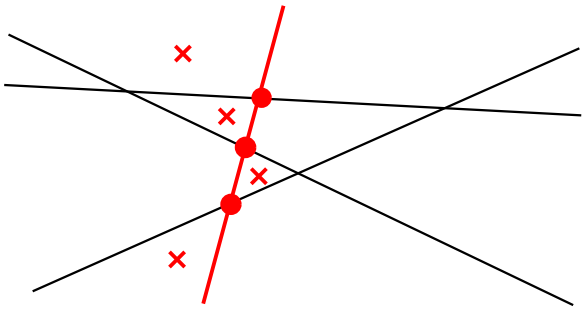


Notice that three of the original regions remain unchanged (Regions 3 and 5 and 7), and the other four are split in two. Do you notice a pattern here? It seems like for every n we've examined, adding the n-th line leaves exactly $n-1$ regions unchanged while the rest are split in two. Let's try to explain why this happens. Remember that we're trying to identify how many regions appear when we draw n lines, so let's assign that value a "name" so we can refer to it; let's say $R(n)$ represents the number of regions created by drawing n lines on the plane so that no two lines are parallel and no intersection point belongs to more than two lines. In these examples we've considered for small values of n, we've looked at what *changes* when we add a new line; that is, we've figured out what $R(n)$ is by already knowing $R(n-1)$. Let's try to adapt our observations so that they apply to any *arbitrary* value of n.

Assume that we know $R(n)$ already. (Why can we do this? Do we know any particular value of $R(n)$ for sure, for some specific n? Which? How?) Say we have an *arbitrary* diagram of n lines on the plane that satisfy the two conditions given in the puzzle statement above. How many regions do these lines create? Yes, exactly $R(n)$. Now, what happens when we add the $(n+1)$-th line? What can we say *for sure* about this line and how it alters the diagram? Well, the only information we really have is that (a) this new line is not parallel to any of the existing n lines and (b) this new line does not intersect any of the already existing intersection points. Now, condition (a) tells us that this new line must intersect *all* of the exisiting n lines; parallel lines don't intersect, and non-parallel lines must intersect somewhere. Thus, we must create n new intersection points on the diagram. Can any of those intersection points coincide with any existing intersection points? No! This is precisely what condition (b) tells us. These two pieces of information together tell us that, no matter how we draw this new line, as long as it satisfies the requirements of the puzzle, we *must* be able to identify n "special" points along that line. Those special points are precisely the points where the new line intersects an existing line.

We'd now like to take these special points and use them to identify new regions in the diagram. Look back to the cases we examined above: identify the new intersection points and see if you can associate them with new regions. Perhaps it would help to label those intersections with a large dot and mark the new regions with an X to make them all stand out. We'll show you one example below, where $n = 4$. What do you notice? Can you use these dots to

help identify how many new regions are created with the addition of that n-th line? Think about this for a minute and then read on.



Exactly! Between any two of the new intersection points, we have a line *segment* that splits a region in two! All that remains is to identify how many new such segments we've created. Since each one corresponds to exactly *one* existing region split in two, this will tell us exactly how many new regions we've created. We've already figured out that this $(n+1)$-th line creates n new intersection points. Think about how these points are arranged on the line. Any two "consecutive" points create a segment, but the extreme points also create infinite segments (that coninue past those extreme points forever). How many are there in total? Exactly $n+1$. (Look at the diagram above, for $n=3$. We see that there are 3 new intersection points and 4 new segments, with two of them being infinite rays.) This means there are $n+1$ line segments that split regions in two, so we have created exactly $n+1$ new regions! This allows us to say that

$$R(n+1) = R(n) + n + 1$$

Phew, what an observation! It took a bit of playing around with examples and making some geometric arguments, but here we are. We've identified some *inductive structure* to this puzzle; we've found how one case depends on another one. That is, we've found how $R(n+1)$ depends on $R(n)$. This hasn't *completely* solved the puzzle, but we are now much closer. All that remains is to replace $R(n)$ with a similar expression, and continually do this until we reach a value we know, $R(1) = 2$. Observe:

$$\begin{aligned}
R(n+1) &= & & & & & & & \cancel{R(n)} & + & n+1 \\
&= & & & & \cancel{R(n-1)} & + & n & + & n+1 \\
&= & & \cancel{R(n-2)} & + & (n-1) & + & n & + & n+1 \\
&\vdots & & & & & & & & \\
&= & & \cancel{R(2)}+3 & + & \cdots & + & n & + & n+1 \\
&= & R(1)+2 & + & 3 & + & \cdots & + & n & + & n+1
\end{aligned}$$

Since we know $R(1) = 2$, we can say

$$R(n+1) = 2 + (2 + 3 + \cdots + n + (n+1)) = 2 + \left(\sum_{k=1}^{n+1} k\right) - 1 = 1 + \sum_{k=1}^{n+1} k$$

and this is a sum we have investigated before! (Also notice that we had to subract 1 because of the missing first term of the sum in parentheses.) Recall that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$, and to represent the sum we have in the equation above, we just replace n with $n+1$. Therefore,

$$R(n+1) = 1 + \frac{(n+1)(n+2)}{2}$$

One final simplification we would like to make is to replace $n+1$ with n throughout the equation, because it makes more sense to have an expression for $R(n)$ (For what values of n is this valid?)

$$R(n) = 1 + \frac{n(n+1)}{2}$$

Finally, we have arrived at an answer to the originally-posed puzzle! In so doing, we employed an *inductive* technique: we explained how one "fact", namely the value of $R(n+1)$, *depends* on the value of a "previous fact", namely $R(n)$, and used these iterative dependencies to work backwards until we reached a particular, *known* value, namely $R(1)$.

We want to point out, again, that the derivation we followed and the observations we made in this section have guided our intuition into an answer, but this has not *rigorously proven* anything. The issue is with the "$\cdots$", which is not a concrete, "officially" mathematical way of capturing the inductive process underlying our technique. Furthermore, our method with the "lines in the plane" problem had us *starting* with a diagram of $n-1$ lines and *building* a new diagram with n lines; is this okay? Why does this actually tell us anything about an *arbitrary* diagram of n lines? Do all such diagrams come from a smaller diagram with one fewer line?

We will, in the next two chapters, learn the necessary tools to fully describe a *rigorous* way of doing what we have done so far, and in the chapter after that, we will employ those tools to make **mathematical induction** officially rigorous. For now, though, we want to give a heuristic definition of induction and continue to examine interesting puzzles and observations that rely on inductive techniques. Practicing with these types of puzzles–learning when to recognize an inductive process, how to work with it, how to use that structure to solve a problem, and so on–will be extremely helpful in the future, and we have no need to delve into technical mathematical detail. (At least, not just yet!)

2.2.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What properties characterize an *inductive* process?

(2) How did we prove that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ is correct? How was our method inductive? (Reread Section 1.4.2 if you forget!)

(3) Why can we take the sum formula mentioned in the previous question and "replace" n with $n+1$ and know that it still holds true? Can we also replace n with $n-1$?

(4) Work through the algebraic steps to obtain our final expression for the sum of the first n squares; that is, verify that

$$\frac{1}{3}(n+1)^3 - \frac{1}{3}(n+1) - \frac{n(n+1)}{2} = \frac{1}{6}n(n+1)(2n+1)$$

(5) Try to recall the argument that adding the $(n+1)$-th line on the plane created *exactly* $n+1$ new regions. Can you work through the argument for a friend and convince him/her that it is valid?

(6) To find the sum of the first n squares, why couldn't we just square the formula for the sum of the first n numbers? Why is that wrong?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Draw 5 lines on the plane (that satisfy the two conditions of the puzzle) and verify that there are 16 regions. Can you also verify that 6 lines yield 22 regions?

(2) Come up with another description of a sequence that goes $1, 2, 3, 4, \ldots, 100$ that is *not* simply all of the numbers from 1 to 100. (Recall the example we gave: all the numbers from 1 to 100 with no "i" in their English spelling.)

(3) Come up with an algebraic expression that relates $(n+1)^4$ to n^4, like we did with cubes.

(**Challenge:** Can you come up with a *geometric* interpretation for the expression you just derived?)

(4) **Challenge:** Let's bump the "lines in the plane" puzzle up one dimension! Think about having n planes in three-dimensional space. How many regions are created? Assume that no two planes are parallel, and no three of them intersect in one line. (Think about how these two conditions are directly analogous to the specified conditions for the "lines" puzzle.)

2.3 Defining Induction

To properly motivate the forthcoming definition of **mathematical induction** as a proof technique, we want to emphasize that the above examples used some intuitive notions of the structure of the puzzle to develop a "solution", where we use quotation marks around *solution* to indicate that we haven't officially proven it yet. In that sense, we ask the following question: What if we had been *given* the formula that we derived and asked to verify it? What if we had not gone through any intuitive steps to derive the formula and someone just told us that it is correct? How could we check their claim? The reason we ask this is because we are really facing that situation now, except the person telling us the formula is ... the very same intuitive argument *we* discovered above!

Pretend you have a skeptical friend who says, "Hey, I heard about this formula for the sum of the first n natural numbers squared. Somebody told me that they add up to $\frac{1}{6}n(n+1)(2n+1)$. I checked the first two natural numbers, and it worked, so it's gotta be right. Pass it on!" Being a logical thinker, but also a good friend, you nod along and say, "I did hear that, but let's make *sure* it's correct for *every* number." How would you proceed? Your friend is right that the first few values "work out" nicely:

$$\begin{aligned} 1^2 = 1 &= \frac{1}{6}(1)(2)(3) \\ 1^2 + 2^2 = 5 &= \frac{1}{6}(2)(3)(5) \\ 1^2 + 2^2 + 3^2 = 14 &= \frac{1}{6}(3)(4)(7) \\ 1^2 + 2^2 + 3^2 + 4^2 = 30 &= \frac{1}{6}(4)(5)(9) \end{aligned}$$

and so on. We could even check, by hand, a large value of n, if we wanted to:

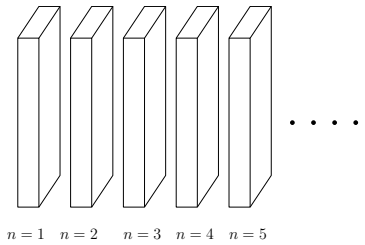
$$1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 = 385 = \frac{1}{6}(10)(11)(21)$$

Remember, though, that this formula is claimed to be valid for *any* value of n. Checking individual results by hand would take forever, because there are an *infinite* number of natural numbers. No matter how many individual values of n we check, there will always be larger values, and how do we *know* that the formula doesn't break down for some large value? We need a far more *efficient* procedure, mathematically and temporally speaking, to somehow verify the formula for all values of n in just a few steps. We have an idea in mind, of course (it's the upcoming rigorous version of mathematical induction), and here we will explain how the procedure works, in a broad sense.

2.3.1 The Domino Analogy

Pretend that we have a set of dominoes. This is a special set of dominoes because we have an infinite number of them (!) and we can imagine anything we want

written on them, instead of the standard array of dots. Let's also pretend that they are set up in an infinite line along an infinite tabletop, and we are viewing the dominos from the side and we can see a label under each one so that we know where we are in the line:



For this particular example, to verify the formula

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

we will imagine a particular "fact" written on each domino. Specifically, we will imagine that the 1st domino has the expression

$$\sum_{k=1}^{1} k^2 = \frac{1}{6}(1)(2)(3)$$

written on it, and the 2nd domino has the expression

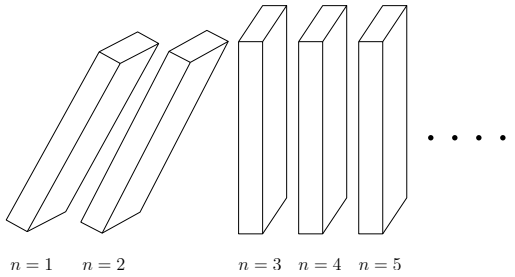
$$\sum_{k=1}^{2} k^2 = \frac{1}{6}(2)(3)(5)$$

written on it. In general, we imagine that the n-th domino in the infinite line has the following "fact" written on it:

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

Since we're dealing with dominos that are meant to fall into each other and knock each other over, let's pretend that whenever a domino falls, that means the corresponding "fact" written on it *is a true statement*. This is how we will relate our physical interpretation of the dominos to the mathematical interpretation of the validity of the formula we derived.

We did check the sum for $n = 1$ by hand: $1^2 = \frac{1}{6}(1)(2)(3)$. Thus, the fact written on the first domino is a true statement, so we know that the first domino will, indeed, fall over. We also checked the sum for $n = 2$ by hand, so we know that the second domino will fall over:



However, continuing like this brings us back to the same problem as before: we don't want to check *every individual* domino to make sure it falls. We would really like to encapsulate our physical notion of the line of dominos—namely, that when a domino falls it will topple into the next one and knock that over, and so on—and somehow relate the "facts" that are written on adjacent dominos.

Let's look at this situation for the first two dominos. Knowing that Domino 1 falls, can we guarantee that Domino 2 falls without rewriting all of the terms of the sum? How are the statements written on the two dominos related? Each statement is a sum of squared natural numbers, and the one on the second domino has exactly one more term. Thus, knowing *already* that Domino 1 has fallen, we can use the *true statement* written on Domino 1 to *verify* the truth of the statement written on Domino 2:

$$\sum_{k=1}^{2} k^2 = 1^2 + 2^2 = 1 + 2^2 = 5 = \frac{1}{6}(2)(3)(5)$$

Now, this may seem a little silly because the only "work" we have saved is not having to "do the arithmetic" to write $1^2 = 1$. Let's use this procedure on a case with larger numbers so we can more convincingly illustrate the benefit of this method. Let's *assume* that Domino 10 has fallen. (In case you are worried about this assumption, we wrote the full sum a few paragraphs ago and you can verify it there.) This means we *know* that

$$\sum_{k=1}^{10} k^2 = \frac{1}{6}(10)(11)(21) = 385$$

is a *true statement*. Let's use this to verify the statement written on Domino 11, which is

$$\sum_{k=1}^{11} k^2 = \frac{1}{6}(11)(12)(23)$$

The sum written on Domino 11 has 11 terms, and the first 10 are exactly the sum written on Domino 10! Since we know something about that sum, let's just separate that 11th term from the sum and apply our knowledge of the other

terms:

$$\begin{aligned}\sum_{k=1}^{11} k^2 &= \left(1^2 + 2^2 + \cdots + 10^2\right) + 11^2 \\ &= \left(\sum_{k=1}^{10} k^2\right) + 11^2 \\ &= 385 + 121 \\ &= 506 \\ &= \frac{1}{6}3036 = \frac{1}{6}(11)(12)(23)\end{aligned}$$

Look at all of the effort we saved! Why bother reading the first 10 terms of the sum if we know something about them already?

Now, imagine if we could do this procedure for *all* values of n, *simultaneously*! That is, imagine that we could prove that any time Domino n falls, we are *guaranteed* that Domino $(n+1)$ falls. What would this tell us? Well, think about the infinite line of dominos again. We *know* Domino 1 will fall, because we checked that value by hand. Then, because we verified the "Domino n knocks over Domino $(n+1)$" step for *all* values of n, we know Domino 1 will fall into Domino 2, which in turn falls into Domino 3, which in turns falls into Domino 4, which ... The entire line of dominos will fall! In essence, we could collapse the whole line of dominos falling down into just *two* steps:

(a) Make sure the first domino topples;

(b) Make sure every domino knocks over the one immediately after it in line.

With only these two steps, we can *guarantee* every domino falls and, therefore, *prove* that all of the facts written on them are true. This will prove that the formula we derived is valid for *every* natural number n.

We have already accomplished step (a), so now we have to complete step (b). We have done this for specific cases in the previous paragraphs (Domino 1 topples Domino 2, and Domino 10 topples Domino 11), so let's try to follow along with the steps of those cases and generalize to an arbitrary value of n. We *assume*, for some *specific but arbitrary* value of n, that Domino n falls, which tells us that the equation

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

is a *true statement*. Now, we want to relate this to the statement written on Domino $(n+1)$ and apply the knowledge given in the equation above. Let's do what we did before and write a sum of $n+1$ terms as a sum of n terms plus the last term:

$$\sum_{k=1}^{n+1} k^2 = 1^2 + 2^2 + \cdots + n^2 + (n+1)^2 = \left(\sum_{k=1}^{n} k^2\right) + (n+1)^2$$

Next, we can apply our assumption that Domino n has fallen (which tells us that the fact written on it is true) and write

$$\sum_{k=1}^{n+1} k^2 = \frac{1}{6}n(n+1)(2n+1) + (n+1)^2$$

Is this the same as the fact written on Domino $(n+1)$? Let's look at what that is, first, and then compare. The "fact" on Domino $(n+1)$ is similar to the fact on Domino n, except everywhere we see "n" we replace it with "$n+1$":

$$\sum_{k=1}^{n+1} k^2 = \frac{1}{6}(n+1)((n+1)+1)(2(n+1)+1) = \frac{1}{6}(n+1)(n+2)(2n+3)$$

It is not clear yet whether the expression we have derived thus far is actually equal to this. We could attempt to simplify the expression we've derived and factor it to make it "look like" this new expression, but it might be easier to just expand both expressions and compare all the terms. (This is motivated by the general idea that expanding a factored polynomial is far easier than recognizing a polynomial can be factored.) For the first expression, we get

$$\begin{aligned}\frac{1}{6}n(n+1)(2n+1) + (n+1)^2 &= \frac{1}{6}n(2n^2+3n+1) + (n^2+2n+1)\\ &= \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n + n^2 + 2n + 1\\ &= \frac{1}{3}n^3 + \frac{3}{2}n^2 + \frac{13}{6}n + 1\end{aligned}$$

and for the second expression, we get

$$\begin{aligned}\frac{1}{6}(n+1)(n+2)(2n+3) &= \frac{1}{6}(n+1)(2n^2+7n+6)\\ &= \frac{1}{6}\left[(2n^3+7n^2+6n) + (2n^2+7n+6)\right]\\ &= \frac{1}{3}n^3 + \frac{3}{2}n^2 + \frac{13}{6}n + 1\end{aligned}$$

Look at that; they're identical! Also, notice how much easier this was than trying to rearrange one of the expressions and "morph" it into the other. We proved they were identical by manipulating them both and finding the same expression, ultimately. Now, let's look back and assess what we have accomplished:

1. We likened proving the validity of the formula

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

for *all* values of n to knocking over an infinite line of dominos.

2. We verified that Domino 1 will fall by checking the formula corresponding to that case by hand.

3. We proved that Domino n will fall into Domino $(n+1)$ and knock it over by *assuming* the fact written on Domino n is true and using that knowledge to show that the fact written on Domino $(n+1)$ must also be true.

4. This guarantees that *all* dominos will fall, so the formula is true for all values of n!

Are you convinced by this technique? Do you think we've *rigorously proven* that the formula is valid for all natural numbers n? What if there were a value of n for which the formula didn't hold? What would that mean in terms of our domino scheme?

Remember that this domino analogy is a good intuitive guide for how induction works, but it is not built on mathematically rigorous foundations. That will be the goal of the next couple of chapters. For now, let's revisit the other example we've examined in this section: lines in the plane. Again, the use of *ellipses* in our derivation of the formula $R(n)$ is troublesome and we want to avoid it. Let's try to follow along with the domino scheme in the context of this puzzle.

Imagine that we have defined the expression $R(n)$ to represent the number of distinct regions in the plane created by n lines, where no two lines are parallel and no three intersect at one point. Also, imagine that on Domino n we have written the "fact" that "$R(n) = 1 + \frac{n(n+1)}{2}$". Can we follow the same steps as above and verify that all the dominos will fall?

First, we need to check that Domino 1 does, indeed fall. This amounts to verifying the statement: "$R(1) = 1 + \frac{1(2)}{2} = 1+1 = 2$". Is this a true statement? Yes, of course, we saw this before; one line divides the plane into two regions. Second, we need to prove that Domino n will topple into Domino $(n+1)$ for any *arbitrary* value of n. That is, let's *assume* that "$R(n) = 1 + \frac{n(n+1)}{2}$" is a *true* statement for some value of n and *show* that "$R(n+1) = 1 + \frac{(n+1)(n+2)}{2}$" must also be a true statement. How can we do this? Well, let's follow along with the argument we used before to relate $R(n+1)$ to $R(n)$. By considering the geometric consequences of adding an extra line to *any* diagram with n lines (that also fit our rules about the lines) we proved that $R(n+1) = R(n)+n+1$. Using this knowledge *and* our assumption about Domino n falling, we can say that

$$R(n+1) = R(n) + n + 1 = 1 + \frac{n(n+1)}{2} + n + 1$$

Is this the same expression as what is written on Domino $(n+1)$? Again, let's simplify *both* expressions to verify they are the same. We have

$$1 + \frac{n(n+1)}{2} + n + 1 = 2 + n + \frac{n^2+n}{2} = \frac{1}{2}n^2 + \frac{3}{2}n + 2$$

and

$$1 + \frac{(n+1)(n+2)}{2} = 1 + \frac{n^2+3n+2}{2} = \frac{1}{2}n^2 + \frac{3}{2}n + 2$$

Look at that; they're identical! Thus, we have shown that Domino n is *guaranteed* to fall into Domino $(n+1)$, for *any* value of n. Accordingly, we can declare that *all* dominos will fall!

Think about the differences between what we have done with this "domino technique" and what we did before to derive the expressions we just proved. Did we use any ellipses in this section? Why is it better to prove a formula this way? Could we have used the domino induction technique to *derive* the formulas themselves?

2.3.2 Other Analogies

The Domino Analogy is quite popular, but it's not the only description of how induction works. Depending on what you read, or who you talk to, you might learn of a different analogy, or some other kind of description altogether. Here, we'll describe a couple that we've heard of before. It will help solidify your understanding of induction (at least as far as we've developed it) to think about how these analogies are all *equivalent*, fundamentally.

Mojo the Magical, Mathematical Monkey

Imagine an infinite ladder, heading straight upwards into the sky. There are infinitely-many rungs on this ladder, numbered in order: 1, 2, 3, and so on going upwards. Our friend Mojo happens to be standing next to this ladder. He is an intelligent monkey, very interested in mathematics but also a little bit magical, because he can actually climb up this infinite ladder!

If Mojo makes it to a certain rung on the ladder, that means the fact corresponding to that number is True. How can we make sure he climbs up the entire ladder? It would be inefficient to check each rung individually. Imagine that: we would have to stand on the ground and make sure he got to Rung 1, then we would have to look up a bit and make sure he got to Rung 2, and then Rung 3, and so on ... Instead, let's just confirm two details with Mojo before he starts climbing. Is he going to start climbing? That is, is he going to make it to Rung 1? If so, great! Also, are the rungs close enough together so that he can *always* reach the next one, no matter where he is? If so, even greater! These are exactly like the conditions established in our Domino Analogy. To ensure that Mojo gets to *every* rung, we just need to know he gets to the first one and that he can always get to the next one.

Doug the Induction Duck

Meet Doug. He's a duck. He also loves bread, and he's going to go searching through everyone's yards to find more bread. These yards are all along Induction Street in Math Town, where the houses are numbered 1, 2, 3, and so on, all in a row.

Doug starts in the yard of house #1, looking for bread. He doesn't find any, so he's still hungry. Where else can he look? The house next door, #2, has a

backyard, too! Doug heads that way, his tummy rumbling. He doesn't find any bread there, either, so he has to keep looking. He already knows house #1 has no bread, so the only place to go is next door to house #3. We think you see where this is going ...

If we were keeping track of Doug's progress, we might wonder whether he eventually gets to *every* yard. Let's say we also knew ahead of time that *nobody* has any bread. This means that whenever he's in someone's yard, he will definitely go to the next house, still searching for a meal. This means that he will definitely get to every house! That is, no matter which house we live in, no matter how large the number on our front door might be, at some point we will see Doug wandering around our backyard. (Unfortunately, he will go hungry all this time, though! Poor Doug.)

2.3.3 Summary

Let's reconsider what we've accomplished with the two example puzzles we've seen thus far, and the analogies we've given. In our initial consideration of each puzzle, we identified some aspect of the *structure* of the puzzle where a "fact" depended on a "previous fact". In the case of the cubic numbers, we found a way to express $(n+1)^3$ in *terms* of n^3; in the case of the lines in the plane, we described how many regions were added when an extra line was added to a diagram with n lines. From these observations, we applied this encapsulated knowledge over and over until we arrived at a "fact" that we knew, for a "small" value of n (in both cases, here, $n = 1$). This allowed us to derive a formula or equation or expression for a general fact that should hold for *any* value of n.

This work was interesting and essential for deriving these expressions, but it was *not enough* to *prove* the validity of the expressions. In doing the work described above, we identified the presence of an inductive process and utilized its structure to derive the expressions in question. This was beneficial in two ways, really; we actually found the expressions we wanted to prove and, by recognizing the inductive behavior of the puzzle, we realized that proving the expressions by *mathematical induction* would be prudent and efficient.

For the actual "proof by induction", we followed two main steps. First, we identified a "starting value" for which we could check the formula/equation by hand. Second, we *assumed* that one particular value of n made the corresponding formula hold true, and then used this knowledge to show that the corresponding formula for the value $n+1$ must also hold true. Between those two steps, we could rest assured that "all dominos will fall" and, therefore, the formulas would hold true for all relevant values of n.

One Concern: What's at the "top" of the ladder?

You might be worried about something, and we'll try to anticipate your question here. (We only bring this up because it's a not uncommon observation to make. If you *weren't* thinking about this, try to imagine where the idea would come from.) You might say, "Hey now, I think I see how Mojo is climbing the ladder,

but how can he actually get *all the way to the top*? It's an infinite ladder, right? And he never gets there . . . right?"

In a way, you would be right. Since this magical ladder really does go on *forever*, then there is truly no *end* to it and Mojo will never get "there". However, that isn't the point; we don't care about any "end" of the ladder (and not just because there *isn't* one). We just need to know that Mojo actually gets to *every possible* rung. He doesn't have to surpass all of them and stand at the top of the ladder, looking down at where he came from. That wasn't the goal!

Think of it this way: pretend you have a vested interest in some particular fact that we're proving. Let's say it's Fact #18,458,789,572,311,000,574,003. (Some huge number. It doesn't matter, really.) Its corresponding rung is waaayyyyyy up there on the ladder, and all you care about is whether or not Mojo gets there on his journey. Does he? . . . You bet he does! It might take a long time (how many steps will it take?), but in this magical world of monkeys and ladders, who cares about time anyway! You know that he'll eventually get there, and that makes you happy. Now, just imagine that for each fact, there's somebody out there in that magical world that cares about only that fact. Surely, everyone will be happy with the knowledge that Mojo will get to their rung on his journey. Nobody cares about whether he gets to the top; that isn't their concern. Meanwhile, out here in our regular, non-magical world, we are extremely happy with the fact that everyone in *that* world will eventually be happy. That entire infinite process of ladder-climbing was condensed into just two steps, and with only those two steps, we can rest assured that every rung on that ladder will be touched. Every numbered fact is true.

Think about this in terms of the Domino Analogy, as well. Do we care whether or not there is some "ending point" of the line of dominoes, so that they all fall into a wall somewhere? Of course not; the line goes on forever. Every domino will eventually fall over, and we don't even care how "long" that takes. Likewise, we know Doug will get to everyone's yard; we don't care "when" he gets to any *individual* yard, just that he gets to *all* of them.

2.3.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How are the Domino, Mojo, and Doug analogies all *equivalent*? Can you come up with some "function" that describes their relationship, that converts one analogy into another?

(2) Find a friend who hasn't studied mathematical induction before, and try to describe it. Do you find yourself using one of the analogies? Was it helpful?

(3) Why is it the case that our work with the cubes didn't *prove* the summation formula? Why did we still need to go through all that work?

(4) Think about the Domino Analogy. Is it a problem that the line of dominoes goes on forever? Does this mean that there are some dominoes that will never fall down? Try to describe what this means in terms of the analogy.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Work through the inductive steps to prove the formula

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

(2) Work through the inductive steps to prove the formula

$$\sum_{k=1}^{n} (2k-1) = n^2$$

(3) Work through the inductive steps to prove the formula

$$\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$$

(4) Suppose we have a series of facts that are indexed by natural numbers. Let's use the expression "$P(n)$" to represent the n-th fact.

   (a) If we want to prove *every* instance is True, for every natural number n, how can we do this?

   (b) What if we want to prove that only every *even* value of n makes a True statement? Can we do this? Can you come up with a modification of one of the analogies we gave that would describe your method?

   (c) What if we want to prove that only every value of n greater than or equal to 4 makes a True statement? Can we do this? Can you come up with a modification of one of the analogies we gave that would describe your method?

2.4 Two More (Different) Examples

This short section serves a few purposes. For one, we don't want you to get the idea, right away, that induction is all about proving a *numerical formula* with numbers and polynomials. Induction is so much more useful than that! One of the following examples, in particular, will be about proving some abstract property is true for any "size" of the given situation. You will see how it still falls under the umbrella of "induction", but you will also notice how it is different from the previous examples. Furthermore, these examples will illustrate that sometimes we need to know "more information" to knock over some dominoes. In the previous examples, we only needed to know that Domino n fell to *guarantee* that Domino $n+1$ will fall. Here, though, we might have to know about several previous dominoes. After these two examples, we will summarize how this differs from the domino definition given above, and preview a broader definition of the technique of induction, as it applies to these examples.

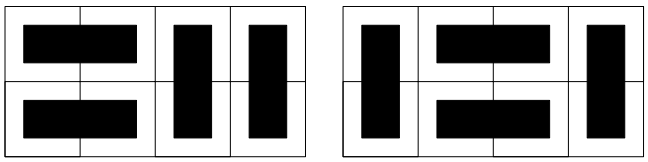
2.4.1 Dominos and Tilings

This next example is a little more complicated than the first two. We will still end up proving a certain numerical *formula*, but the problem is decidedly more visual than just manipulating algebraic expressions. Furthermore, we'll notice an interesting "kink" in the starting steps, where we have to solve a couple of "small cases" before being able to generalize our approach. This will be our first consideration of how the technique of induction can be generalized and adapted to other situations.

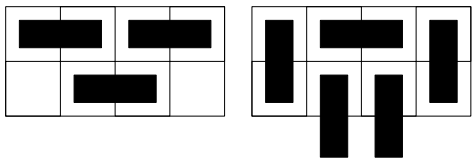
The question we want to answer is nicely stated as follows:

> Given a $2 \times n$ array of squares, how many different ways can we tile the array with dominoes? A *tiling* must have every square covered by one–and *only* one–domino.

For example, the following are proper tilings

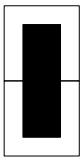


whereas the following are *not* proper tilings

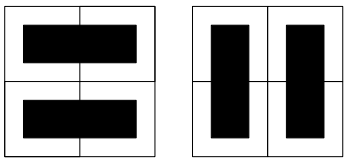


As before, let's examine the first few cases–where $n = 1, 2, 3$, and so on–and see if we notice any patterns. Try working with the problem yourself, before reading on, even!

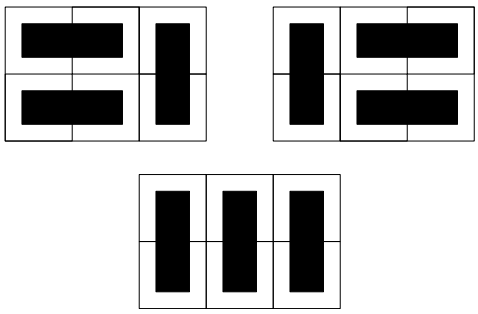
When $n = 1$, we have an array that is exactly the shape of one domino, so surely there is only one way to do this. Let's use the notation $T(n)$ to represent the number of tilings on a $2 \times n$ array. Thus, $T(1) = 1$.



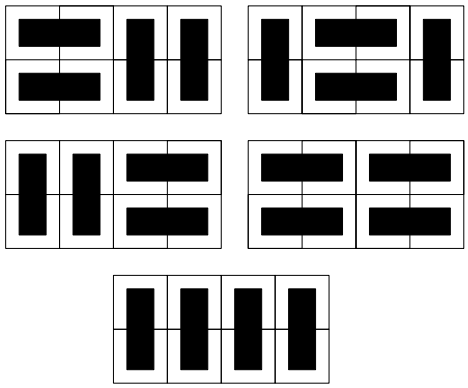
When $n = 2$, we have a 2×2 array. Since the orientation of the array matters, we have each of the following distinct tilings. Thus, $T(2) = 2$.



What about when $n = 3$? Again, we can enumerate these tilings by hand and be sure that we aren't missing any. We see that $T(3) = 3$.

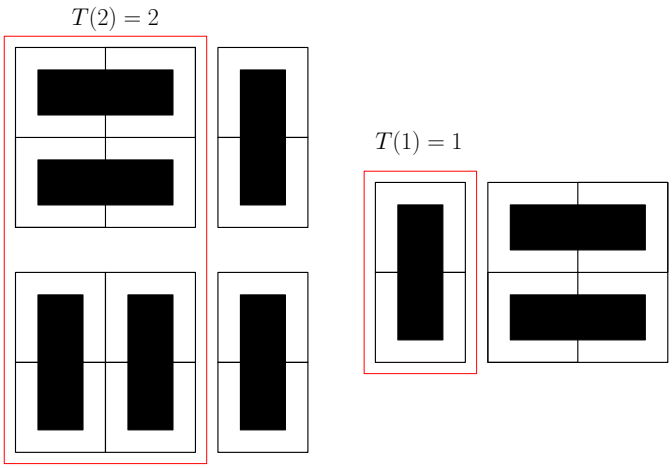


Okay, one more case, when $n = 4$. We see that $T(4) = 5$.



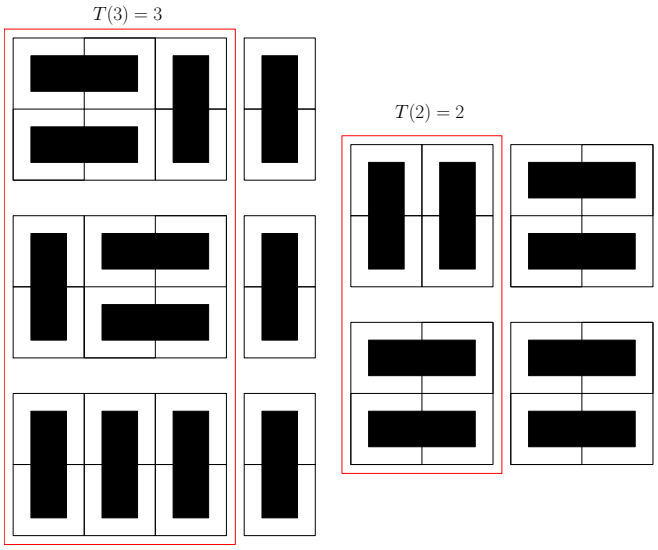
Can we start to find a pattern now? Writing out larger arrays will just be tiresome! Let's think about how we could have used the fact that $T(1) = 1$ to deduce something about $T(2)$ ... Well, wait a minute ... We couldn't, right? There was something fundamentally different about those two cases. Specifically, because dominoes are 2×1 in size, the fact that we only added one row to the array didn't help us.

Alright, let's consider $n = 3$, then. Could we use the fact that $T(2) = 2$ at all? In this case, yes! Knowing there were two tilings of the 2×2 array, we could immediately build two tilings of the 2×3 array without much thought. Specifically, we can just *append a vertical domino* to each of those previous tilings. But we know now that $T(3) = 3$. Where did the third tiling come from? Look at that tiling again and how it compares to the other two. In that third tiling, the dominoes on the right side are horizontal, as opposed to the vertical one in the other two tilings. If we remove those two parallel, horizontal dominoes, we are left with precisely the situation when $n = 1$. Put another way, we can build a tiling of a 2×3 array by *appending a square of two horizontal dominoes* to the right side. In total, then, we have described all of the tilings of a 2×3 board in terms of boards of smaller sizes, namely 2×2 and 2×1:



$$T(3) = 3 = 2 + 1 = T(2) + T(1)$$

Now you might see how the pattern develops! Let's show you what happens when $n = 4$, how we can construct *all* of the tillings that make up $T(4)$ by appending a vertical domino to each of the tilings that make up $T(3)$, or by appending two horizontal dominoes to each of the tilings that make up $T(2)$:



Notice, as well, that no tiling for the 2×4 array was produced *twice* in this way. (Think carefully about why this is true. How can we characterize the two types of tilings in a way that will guarantee they don't coincide at all?) With this information, we can immediately conclude that $T(4) = T(3) + T(2)$.

Furthermore, we can generalize this argument; nothing was special about $n = 4$, right? For any particular n, we can just consider all possible tilings, and look at what happens on the *far right-hand side* of the array: either we have one vertical domino (which means the tiling came from a $2 \times (n-1)$ array) or two horizontal dominoes (which means the tiling came from a $2 \times (n-2)$ array). With confidence in this argument, we can conclude that

$$T(n) = T(n-1) + T(n-2)$$

for all of the values of n for which this expression makes sense. What values are those? Remember that we had to identify $T(1)$ and $T(2)$ separately; this argument doesn't apply to those values. Accordingly, we have to add the restriction $n \geq 3$ for the equation above to hold true.

With this information, we can then easily figure out $T(n)$ for any value of n, given enough time. We could write a computer program fairly easily, even. It was this *inductive* argument, though—the pattern that we noticed and our thorough description of why it occurs—that allowed us to make the conclusion in the first place. In this case, too, it just so happens that the value of every term, $T(n)$, depends on the value of *two* previous terms, $T(n-1)$ and $T(n-2)$. This did *not* happen in our previous examples in this chapter, and it hints at something deeper going on here. Do you see how our previous definition of induction, and the domino analogy, doesn't exactly apply here anymore? How might you try to amend our analogy to explain this kind of situation? Think

about these issues for a bit and then read on. We'll talk about them more in-depth after the next example.

By the way, did you notice something interesting about our solution to this example? Do you know any other sequences of numbers that behave similarly? Think about it ...

2.4.2 Winning Strategies

This example will be our first induction puzzle that *doesn't* prove a numerical formula! It might seem strange to think about that, but it's true, as you'll see. This is actually more common in mathematics than you might think, too: a problem or mathematical object might have some underlying inductive structure without depending on something algebraic or arithmetic.

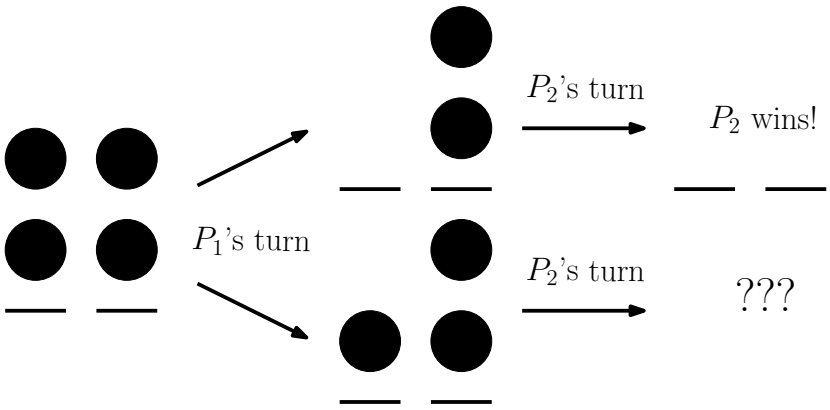
In fact, we will be discussing a *game*. It's a game in the usual sense—there are rules to be followed by two players and there is a clear winner and loser—but it's also a game in the mathematical sense, where we can formulate the rules and playing situations using mathematical notation and discuss formal *strategies* in an abstract way. We can even *solve* the game. This is very different than say, the game of baseball.

Let's discuss the rules for this game, which we shall call "Takeaway", for now. There are two players, called P_1 and P_2. The player P_1 goes first every time. The players start with two piles of stones in the middle of a table, each pile containing exactly n stones, where n is some natural number. (To distinguish the different versions of the game, we will say the players are "playing T_n" when there are n stones per pile.) On each player's move, they are allowed to remove *any* number of stones from *either* pile. It is illegal, though, to remove stones from both piles at once. The player who removes the *final* stone from the piles is the *winner*.

Try playing Takeaway with some friends. Use pennies or candies or penny candy as stones. Try it for different values of n. Try switching roles so you are P_1 and then P_2. Try to come up with a winning *strategy*, a method of playing that maximizes your chances of winning. Try to make a conjecture for what happens for different values of n. Who is "supposed" to win? Can you *prove* your claim? Seriously, play around with this game and attempt to prove something before reading on for our analysis thereof. You might be surprised by what you can accomplish!

As with the other examples, let's use some small values of n to figure out what's really going on, then try to generalize. When $n = 1$, this game is rather silly. P_1 must empty one pile of its only stone, then P_2 gets the only remaining stone in the other pile. Thus, P_2 wins. (Notice that it doesn't matter which of the two piles P_1 picks from, P_2 will always get the other one. We might say that P_1 picks the pile on the left "without loss of generality" because it doesn't matter either way; the situations are equivalent, so we might as well say it's the left pile to have something concrete to say. We will explore this idea of "without loss of generality" later on when we discuss mathematical logic, too.)

When $n = 2$, we now have a few cases that might appear. Think about P_1's possible moves. Again, they might act on either the left or right pile, but because they're ultimately identical and we can switch the two piles, let's just say (without loss of generality) that P_1 removes some stones from the left pile. How many? It could be one or two stones. Let's examine each case separately.

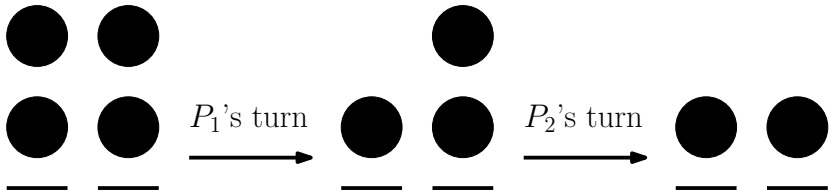


If P_1 removes both stones, how should P_2 react? Duh, they should take the other pile, so P_1 probably shouldn't have made that move in the first place. However, P_1 might not be thinking straight or something and, besides, we need to consider all possible situations here to fully analyze this game. Thus, in this case (the top line in the above diagram) P_2 wins. Okay, that's the easy situation.

What if P_1 removes just one stone from the left pile (the bottom line above)? How should P_2 react? We now have some options:

- If P_2 removes the other stone from the left pile ... well, P_1 takes the other pile and P_1 wins.
- If P_2 removes both stones from the right pile ... well, P_1 takes the last stone from the left pile and P_1 wins.

- However, if P_2 removes just one stone from the right pile ...



Now we have exactly the same situation presented by T_1, which we already analyzed! It is, again, P_1's move first, so we know what will happen: P_2 wins no matter what. If you are player P_2, this is obviously the best move: you win *no matter how* P_1 *responds*!

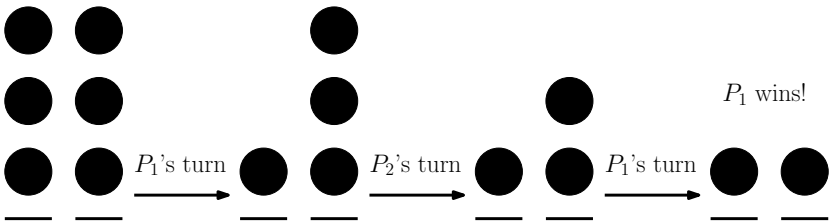
Stepping back for a second, let's think about what this has shown: no matter what P_1 does first (remove one or two stones from either pile), there is *some possible response* that P_2 can make that will *guarantee* a win for P_2, regardless of P_1's subsequent response. Wow, P_2 is sitting pretty! Let's see if this happens for other values of n.

When $n = 3$, we will again assume (without loss of generality) that player P_1 acted on the left pile. They could remove one, two, or three stones:

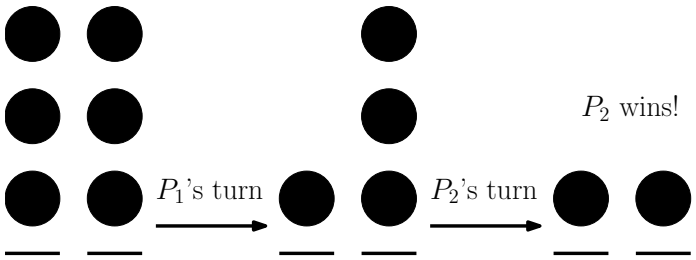
- If P_1 removes all three, P_2 responds by taking the other pile completely and wins.

- If P_2 removes two stones ... well, what should player P_2 do?

Finishing off that left pile is stupid (because P_1 can take the whole right pile and win), and pulling the entire right pile is similarly stupid (because P_1 can take the whole left pile and win), so something in between is required. Now, if P_2 removes just one stone from the right pile, notice that P_1 can respond with the same action; this leaves exactly one stone in both piles, but the roles reversed. With P_2 going first in such a situation, they are now bound to lose, per our previous analysis. Bad move, P_2!



Let's try again. If P_2 removes two stones from the right pile instead ... look at that! We now have exactly one stone in each pile, with P_1 up first, so we know P_1 is going to lose. P_2 strikes again!



Think about the case where $n = 4$ for a minute, and you'll find the exact same analysis occurring. You'll another possibility to consider: player P_1 can remove one, or two, or three, or four stones from the left pile. Whatever they do, though, you'll find that P_2 can just *mimic that action* on the other pile, reducing the whole game to a previous, *smaller* version of the game, where P_2 was shown to be guaranteed a win! It looks like P_2 is in the driver's seat the whole time, since they can respond to whatever P_1 does, making an identical move on the other pile. No matter what P_1 does, there is always a response for P_2 that means they win, regardless of P_1's subsequent moves. In this sense, we say "P_2 has a winning strategy". There is a clear and describable method for P_2 to assess the game situation and choose a specific move to *guarantee a win.*

How might we prove this? How does this even fit into this chapter on induction? It might be hard to see, at the moment. What are we really even proving here? What are the dominoes or rungs in our analogy for this problem? In wrapping your brain around this example, you should hopefully realize the following: induction is *not* about algebraic formulas all the time; induction represents some kind of "building-up" structure, where larger situations depend on smaller ones; we have to prove some initial fact, and then argue how an arbitrary, larger fact can be reduced so that it depends on a previous fact. This is really what the dominoes analogy is meant to accomplish. It just so happens that this analogy is particularly illustrative for certain induction problems (but not all) and is visualizable and memorable. It does not perfectly apply to *all* situations, though.

Read back through these four examples from this chapter and think about how they are similar and how they are different. Try to come up with a more precise, mathematical description of mathematical induction using some better terminology, perhaps of your own invention. (By this, we mean something better than our intuitive analogy. You'd be surprised at how well you might be able to describe induction without really knowing what you "ought" to say, and you'll actually learn a lot, in the process!) In due time, we will have a rigorous statement to make, and prove, about mathematical induction and its various forms. In the meantime, we need to take a trip through some other areas of mathematics to build up the necessary language, notation, and knowledge to come back and tackle this problem. Before we go, though, we should mention a few useful applications of mathematical induction.

2.4.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How are these two examples *inductive*? In what ways are they similar to the previous examples, with the cubes and lines? In what ways are they different?

(2) With the domino tilings, how *many* previous values did we need to know to compute $T(n)$?

(3) What is the difference between writing $T(n) = T(n-1) + T(n-2)$ and $T(n+2) = T(n+1) + T(n)$?

(4) What is the winning strategy in the Takeaway game? Try playing with a friend who doesn't know the game, and use that strategy as player P_2. How frustrated do they get every time you win? Do they start to catch on?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) What is $T(5)$? Can you draw all of those tilings?

(2) Work through the possibilities for takeaway with two piles of 4 stones. Can you make sure that player P_2 always has a winning move?

(3) **Challenge:** What happens if you play Takeaway with *three* piles of equal sizes? Can you find a winning strategy for either player? Try playing with a friend and see what happens!

(4) Look up the *Fibonacci numbers*. How are they related to the sequence of numbers $T(n)$ we found in the domino tiling example?

2.5 Applications

2.5.1 Recursive Programming

The concepts behind mathematical induction are employed heavily in computer science, as well. Think back to how we first derived the formula for $\sum_{k=1}^{n} k^2$.

Once we had a way to represent a cubic number in terms of a smaller cube and some leftover terms, we repeated this subsitution process over and over until we arrived at the "simplest" case, namely, the one that we first observed when starting the problem: $2^3 = 1+3+3+1$. Recursive programming takes advantage of this technique: to solve a "large" problem, identify how the problem depends on "smaller" cases, and reduce the problem until one reaches a simple, known case.

A classical example of this type of technique is seen in writing code to compute the *factorial* function, $n!$, which is defined as the product of the first n natural numbers:

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$$

This is a simple definition that we, as humans, intuitively understand, but telling a computer how to perform this product doesn't work quite the same way. (Try it! How do you say "and just keep going until you reach n" in computer code?) A more efficient way to program the function, and one that models the mathematically inductive definition, in fact, is to have one program *recursively call itself* until it reaches that "simple" case. With the factorial function, that case is $1! = 1$. For any other value of n, we can simply apply the knowledge that

$$n! = (n-1)! \cdot n$$

over and over to compute $n!$. Consider the following *pseudocode* that represents this idea:

```
 factorial(n):

if n = 1
   return 1
else
   return n * factorial(n-1)
end
```

We know that $1! = 1$, so if the program is asked to compute that, the correct value is returned right away. For any larger value of n, the program refers to *itself* and says, "Go back and compute $(n-1)!$ for me, then I'll add a factor of n at the end, and we'll know the answer." To compute $(n-1)!$, the program asks, again, if the input is 1; if not, it calls itself and says, "Go back and compute $(n-2)!$ for me, then I'll add a factor of $n-1$ at the end." This process continues until the program returns $1! = 1$. From there, it knows how to find $2! = 1 \times 2$, and then $3! = 2! \times 3$, and so on, until $n! = (n-1)! \times n$.

Another example involving recursive programming arises with the *Fibonacci numbers*. You may have seen this sequence of numbers before in a mathematics course. (In fact, we even mentioned them in the last section, with the domino tilings!) You also might have heard about how they appear in nature in some interesting and strange ways. (The sequence was first "discovered" by the Italian mathematician Leonardo of Pisa while studying the growth of rabbit populations.) The first two numbers in the sequence are specified to be 1, and any

number in the sequence is defined as the sum of the previous two. That is, if we say $F(n)$ represents the n-th Fibonacci number, then

$$F(1) = 1 \quad \text{and} \quad F(2) = 1 \quad \text{and} \quad F(n) = F(n-1) + F(n-2) \text{ for every } n \geq 3$$

Now, what is $F(5)$? Or $F(100)$? Or $F(10000)$? This can be handled quite easily by a recursive program. The idea is the same: if the program refers to either one of the "simple cases", i.e. $F(1)$ or $F(2)$, then it will know to return the correct value of 1 immediately. Otherwise, it will call itself to compute the previous two numbers and then add those together. Look at the pseudocode below and think about how it works. What would happen if we used this program to compute $F(10)$? How would it figure out the answer?

```
 Fibonacci(n):

if n = 1 or n = 2
   return 1
else
   return Fibonacci(n-1) + Fibonacci(n-2)
end
```

This follows the same idea as the `factorial` program above (let the program call itself to compute values for "smaller" cases of the function until we reach a known value) but there's something a little deeper going on here. If we were to input $n = 10$ into the program, it would recognize that it does not know the output value yet, and it will call itself to compute `Fibonacci(9)` and `Fibonacci(8)`. In each of those calls to the program, it would again recognize the value is as yet unknown. Thus, it would call upon itself again to compute `Fibonacci(8)` and `Fibonacci(7)`, but also `Fibonacci(7)` and `Fibonacci(6)`. That's right, the program calls itself multiple times with the *same input value*. To compute $F(9)$, we need to know $F(8)$ and $F(7)$, but meanwhile, to compute $F(8)$, we also need to know $F(7)$ and $F(6)$. In this way, we end up calling the program `Fibonacci` many times.

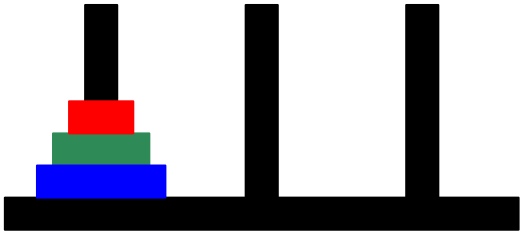
Try to compare the programs `Fibonacci` and `factorial`, especially in regards to the inductive processes we have been investigating in this chapter. Do they use similar ideas? How do they relate to the "domino" analogy of mathematical induction that we outlined? Think of the "fact" written on Domino n as being the computation of the correct value of $n!$ or $F(n)$. How does the analogy work in each case? Will all the dominos fall? Keep these questions in mind as you read on. There is some very powerful mathematics underlying all of these ideas.

2.5.2 The Tower of Hanoi

Let's take a short break and play a game. Well, it's not exactly a break because this is, in a sense, an *inductive* game, so it's completely relevant. But it is a

game, nevertheless! The *Tower of Hanoi* is a very popular puzzle, partly because it involves such simple equipment and rules. Solving it is another matter, though!

Imagine that we have three vertical rods and three disks of three different sizes (colored blue, green, and red) stacked upon each other like so:



The goal is to move all three disks to another rod (either the middle or the right one, it doesn't matter) by following these rules:

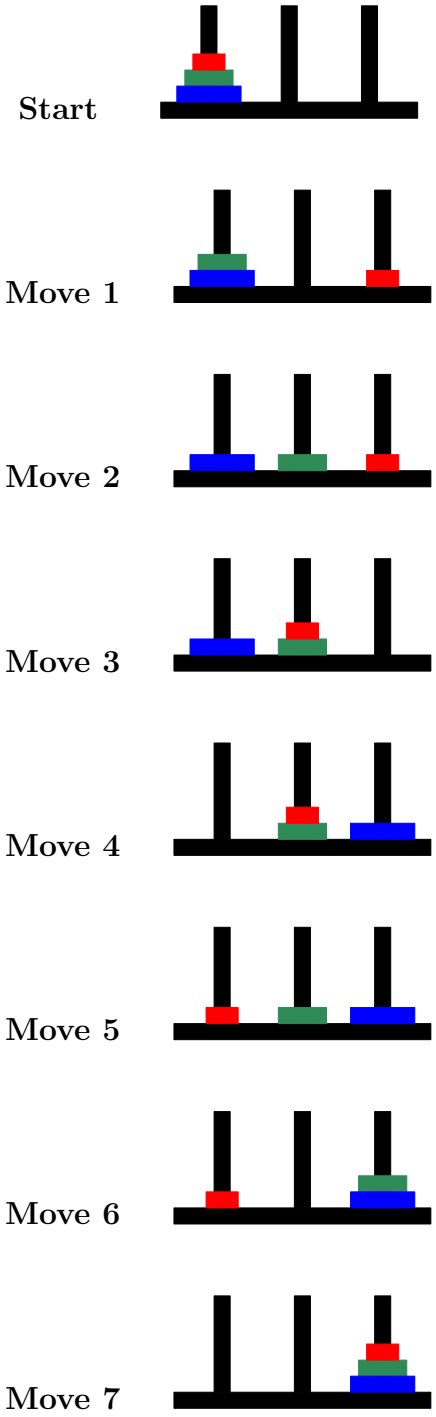
1. A single move consists of moving one (and *only* one) disk from the top of the stack on any rod and moving it to the top of the stack on another rod.

2. A disk cannot be placed on top of a smaller disk.

That's it! Two simple rules, but a difficult game to play. Try modeling the game with a few coins or playing cards or whatever you have handy. (You can even buy Tower of Hanoi sets at some games stores.) Can you solve it? How many moves did it take you? Is your solution the "best" one? Why or why not?

We mentioned that this is an *inductive* game, so let's explore that idea now. We want to consider how many *moves* it takes to solve the puzzle (where one *move* accounts for moving one disk from one rod to another) and, more specifically, identify the *smallest possible number of moves* it would take to solve the puzzle. To solve the puzzle with three disks, we could keep moving the smallest disk back and forth between two rods and generate 100 moves, if we wanted to, and then solve it, but that's certainly not the best way to do it, right? Let's say we found a way to solve the puzzle in a certain number of moves; how could we show that the number of moves we used is the *smallest possible* number of moves?

To address this question, we want to break down the method of solving the puzzle *recursively*. In doing so, we are actually going to answer a far more general question: What is the smallest number of moves required to solve the Tower of Hanoi puzzle with n disks on 3 rods? We posed the puzzle above with just 3 disks to give you a concrete version to think about and work with, but we can answer this more general question by thinking carefully. To make sure we are on the same page, we will show you how we solved the version with 3 disks:


Start
Move 1
Move 2
Move 3
Move 4
Move 5
Move 6
Move 7

Notice that the largest disk is essentially "irrelevant" for most of the solution. Since we are allowed to place any other disk on top of it, all we need to do is "uncover" that disk by moving the other disks onto a different rod, move the largest disk to the only empty rod, then move the other disks on top of the large one. In essence, we perform the same procedure (shifting the two smaller disks from one rod to another) twice and, in between those, we move the large disk from one rod to another. If the largest disk hadn't been there at all, what we actually did was solve the version of the puzzle with 2 disks, but twice! (Think carefully about this and make sure you see why this is true. Follow along with the moves in the diagrams above and pretend the large, blue disk isn't there.)

This shows that the way to solve the 3-disk puzzle involves two iterations of solving the 2-disk puzzle, with one extra move in between (moving the largest disk). This indicates a *recursive* procedure to solve the puzzle, in general. To optimally solve the n-disk puzzle, we would simply follow the procedure to optimally solve the $(n-1)$-disk puzzle, use one move to shift the largest, n-th disk, then solve the $(n-1)$-disk puzzle again.

Now that we have some insight into *how* to optimally solve the puzzle, let's identify how many *moves* that procedure requires. Recognizing that solving this puzzle uses a *recursive* algorithm, we realize that *proving* anything about the optimal solution will require *induction*. Accordingly, we would need to identify a "starting point" for our line of dominos, and it should correspond to the "smallest" or "simplest" version of the puzzle. For the Tower of Hanoi, this is the 1-disk puzzle. Of course, this is hardly a "puzzle" because we can solve it in one move, by simply shifting the only disk from one rod to any other rod. If we let $M(n)$ represent the number of *moves* required to optimally solve the n-disk puzzle, then we've just identified $M(1) = 1$. To identify $M(2)$, we can use our observation from the previous paragraph and say that

$$\underbrace{M(2)}_{\text{solve 2-disk}} = \underbrace{M(1)}_{\text{solve 1-disk}} + \underbrace{1}_{\text{shift largest disk}} + \underbrace{M(1)}_{\text{solve 1-disk}} = 1+1+1 = 3$$

and then it must be that

$$M(3) = M(2) + 1 + M(2) = 3 + 1 + 3 = 7$$

and

$$M(4) = M(3) + 1 + M(3) = 7 + 1 + 7 = 15$$

and so on. Do you notice a pattern yet? Each of these numbers is one less than a power of 2, and specifically, we notice that $M(n) = 2^n - 1$, for each of the cases we have seen thus far. It's important to point out that observing this pattern doesn't *prove* the pattern; just because it works for the first 4 cases does not mean the trend will continue, but that's exactly what an induction proof would accomplish. Also, recognizing that pattern and "observing" that $M(n) = 2^n - 1$ is a non-trivial matter, itself. We happened to know the answer and had no problem identifying the formula for you. You should probably try, on your own, to "solve" the following relationship

$$M(n) = 2M(n-1) + 1 \quad \text{and} \quad M(1) = 1$$

and see if you can derive the formula $M(n) = 2^n - 1$. The reason such a formula is *nicer* than the above relationship is that, now, $M(n)$ depends only on n, and not on previous terms (like $M(n-1)$, for example). This relationship and others like it are known as *recurrence relations*, and they can be rather difficult to solve, in general!

We know how to solve this one, though, and it yields $M(n) = 2^n - 1$. We will leave it to you to verify this. You can do so by checking a few values in the equation above, but we all know that isn't a *proof*. Try working through the inductive steps to actually prove it! We have already done most of the work, but it will be up to you to arrange everything carefully and clearly. Remember that you should identify what the "fact" on each domino is, ensure that Domino 1 falls, and then make a general argument about Domino n toppling into Domino $(n+1)$. Try to write that proof. Do the details make sense to you? Try showing your proof to a friend and see if they understand it. Did you need to tell them anything else or guide them through it? Think about the best way to *explain* your method and steps so that the written version suffices and you don't have to add any verbal explanations.

2.5.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How is a recursive program inductive?

(2) What is the inductive structure of the Tower of Hanoi? Where did we solve the 2-disk puzzle while solving the 3-disk puzzle?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Follow the steps of the psuedocode `factorial` to compute 5!.

(2) Follow the steps of the psuedocode `Fibonacci` to compute $F(5)$.

(3) Solve the Tower of Hanoi puzzle with 4 disks. Make sure that you can do it in the *optimal* number of moves, $2^4 - 1 = 15$.

2.6 Summary

We have now seen some examples of **inductive arguments**. We realized that some of the puzzles we were solving used similar argument styles, and explored several examples to get a flavor for the different issues that might come up in such arguments. Specfically, we saw how inductive arguments are *not* always about proving a summation formula or an equation: inductive arguments can apply to *any* situation where a fact depends on a "previous instance" of that fact. This led us into developing an analogy for how induction works, mathematically speaking. We are comfortbal with thinking of induction in terms of the "Domino Analogy" for now, but one of our main goals in moving forwards is rigorously *stating* and *proving* a principle of induction. For now, let's get lots of practice working with these kinds of arguments. This is what this chapter's exercises are meant to achieve. Later on, once we've formalized induction, we'll be better off for it, and we'll have a thorough understanding of the concept!

2.7 Chapter Exercises

Here are some problems to get you comfortable working with inductive-style arguments. We aren't looking for fully rigorous proofs here, just a good description of what is going on and a write-up of your steps. We'll come back to some of these later and rigorously prove them, once we've established the Principle of Mathematical Induction (PMI) and a corresponding proof strategy.

Problem 2.7.1. Prove the following summation formula holds for every natural number, and for $n = 0$, as well:

$$\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$$

Follow-up question: use this result to state how many games are required to determine a winner in a single-elimination bracket tournament with 2^n teams. (For example, the NCAA March Madness Tournament uses this format, with $n = 6$.)

Problem 2.7.2. Prove that $3^n \geq 2^{n+1}$ for every natural number n that is greater than or equal to 2.

Problem 2.7.3. For which natural numbers n do the following inequalities hold true? State a claim and then prove it.

1. $2^n \geq (n+1)^2$
2. $2^n \geq n!$
3. $3^{n+1} > n^4$
4. $n^3 + (n+1)^3 > (n+2)^3$

Problem 2.7.4. **The December 31 Game:** Two players take turns naming dates from a calendar. On each turn, a player can increase the month or date but not both. The starting position is January 1, and the winner is the person who says December 31. Determine a winning strategy for the first player.

For example, a sequence of moves that yields Player 1 winning is as follows:

> (1) January 10, (2) March 10, (1) August 10, (2) August 25, (1) August 28, (2) November 28, (1) November 30, (2) December 30, (1) December 31

By *winning strategy* we mean a method of play that Player 1 follows that *guarantees* a win, no matter what Player 2 does.

Problem 2.7.5. Find and prove a formula for the sum of a *geometric series*, which is a series of the form

$$\sum_{i=0}^{n-1} q^i$$

for some real number q and some natural number n. (Hint: be careful when $q = 1$.)

Problem 2.7.6. Write a sentence that depends on n such that the sentence is true for all values of n from 1 to 99 (inclusive), but such that the sentence is false when $n = 100$.

Problem 2.7.7. What is wrong with the following "spoof" of the claim that $a^n = 1$ for every n?

"Spoof": Let a be a nonzero real number. Notice that $a^0 = 1$. Also, notice that we can inductively write

$$a^{n+1} = a^n \cdot a = a^n \cdot \frac{a^n}{a^{n-1}} = 1 \cdot \frac{1}{1} = 1$$

"□"

Problem 2.7.8. In a futuristic society, there are only two different denominations of currency: a coin worth 3 Brendans, and a coin worth 8 Brendans. There is also a nation-wide law that says shopkeepers can only charge prices that can be paid in **exact change** using these two coins.

What are the legal costs that a shopkeeper could charge you for a cup of coffee?

Hint: Try a bunch of small values and see what happens.

Problem 2.7.9. Consider a chessboard of size $2^n \times 2^n$, for some arbitrary natural number n. Remove **any** square from the board. Is it possible to tile the remaining squares with L-shaped triominoes?

If your answer is Yes, prove it.

If your answer is No, provide a counterexample argument. (That is, find an n such that no *possible* way of tiling the board will work, and show why this is the case.)

Problem 2.7.10. Consider an $n \times n$ grid of squares. How many sub-squares, of any size, exist within this grid? For example, when $n = 2$, the answer is 5: there are 4 1×1 squares and 1 2×2 square. Find a formula for your answer and try to prove it is correct.

Problem 2.7.11. Prove that, in a line of at least 2 people, if the 1st person is a woman and the last person is a man, then somewhere in the line there is a man standing immediately behind a woman.

Problem 2.7.12. Prove that $n^3 - n$ is a multiple of 3, for every natural number n.

Problem 2.7.13. A **binary n-tuple** is an ordered string of 0s and 1s, with n total numbers in the string. Provide an *inductive argument* to explain why there are 2^n possible binary n-tuples.

Problem 2.7.14. Recall that the **Fibonacci Numbers** are defined by setting $f_0 = 0$ and $f_1 = 1$ and then, for every $n \geq 2$, setting $f_n = f_{n-1} + f_{n-2}$. This produces the sequence $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$

You might not know that the Fibonacci Numbers also have a *closed form*; that is, there is a specific *formula* that defines them, in addition to the usual recursive definition given above. Here it is:

$$f_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right]$$

Prove that this formula is correct for all values of $n \geq 0$.

Problem 2.7.15. Again, considering the Fibonacci Numbers, f_n, prove the following:

1. $\sum_{i=0}^{n} f_i = f_{n+2} - 1$
2. $\sum_{i=0}^{n} f_i^2 = f_n \cdot f_{n+1}$
3. $f_{n-1} \cdot f_{n+1} - f_n^2 = (-1)^n$
4. $f_{m+n} = f_n \cdot f_{n+1} + f_{m-1} \cdot f_n$
5. $f_n^2 + f_{n+1}^2 = f_{2n+1}$

Problem 2.7.16. Try to provide an inductive argument that explains why every natural number $n \geq 2$ can be written as a product of prime numbers. Can you also show that this product is *unique*? That is, can you also explain why there is *exactly one way* to factor a natural number into primes?

Problem 2.7.17. Prove that

$$\sum_{k=1}^{n} k \cdot k! = 1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n! = (n+1)! - 1$$

Problem 2.7.18. What is wrong with the following "spoof" that all pens have the same color.

"Spoof": Consider a group of pens with size 1. Since there is only 1 pen, it certainly has the same color as itself.

Assume that any group of n pens has only one color represented inside the group. (Note: we explained why this assumption is valid for $n = 1$ already, so we can make this assumption.) Take any group of $n + 1$ pens. Line them up on a table and number them from 1 to $n + 1$, left to right. Look at the first n of them, i.e. look at pens $1, 2, 3, \ldots, n$. This is a group of n pens so, by assumption, there is only one color represented in the group. (We don't know what color that is yet.) Then, look at the last n of the pens; i.e. look at pens $2, 3, \ldots, n + 1$. This is also a group of n pens so, by assumption, there is only one color represented in this group, too. Now, pen #2 happens to belong to both of these groups. Thus, whatever color pen #2 is, that is also the color of every pen in *both* groups. Thus, all $n + 1$ pens have the same color.

By induction, this shows that any group of pens, of any size, has only one color represented. Looking at the finite collection of pens in the world, then, we should only find one color. "□"

Problem 2.7.19. ⋆ This problem is *extremely difficult* to analyze, and is taken from the blog of the famous mathematician Terence Tao (link here).

There is an island upon which a tribe resides. The tribe consists of 1000 people, with various eye colours. Yet, their religion forbids them to know their own eye color, or even to discuss the topic; thus, each resident can (and does) see the eye colors of all other residents, but has no way of discovering his or her own (there are no reflective surfaces). If a tribesperson does discover his or her own eye color, then their religion compels them to commit ritual suicide at noon the following day in the village square for all to witness. All the tribespeople are highly logical and devout, and they all know that each other is also highly logical and devout (and they all know that they all know that each other is highly logical and devout, and so forth).

(For the purposes of this logic puzzle, "highly logical" means that any conclusion that can logically deduced from the information and observations available to an islander, will automatically be known to that islander.)

Of the 1000 islanders, it turns out that 100 of them have blue eyes and 900 of them have brown eyes, although the islanders are not initially aware of these statistics (each of them can of course only see 999 of the 1000 tribespeople).

One day, a blue-eyed foreigner visits to the island and wins the complete trust of the tribe.

One evening, he addresses the entire tribe to thank them for their hospitality.

However, not knowing the customs, the foreigner makes the mistake of mentioning eye color in his address, remarking *how unusual it is to see another blue-eyed person like myself in this region of the world.*

What effect, if anything, does this faux pas have on the tribe?

2.8 Lookahead

In this chapter, we have introduced you to the concept of **mathematical induction**. We looked at a few examples of puzzles where an inductive process guided our solution, and then we described how a *proof by induction* would follow to *rigorously verify* that solution. With the mathematical techniques and concepts we have at hand thus far, we had to rely on a non-technical analogy to describe this process to you. Thinking of an infinite line of dominos with "facts" written on them knocking into each other is a perfectly reasonable interpretation of this process, but it fails to represent the full mathematical extent of induction. In a way, it's like having a friend describe to you how to swing a golf club, even though you've never played golf before. Sure, they can provide you with some mental imagery of what a swing "feels like", but without getting out there and practicing yourself, how will you truly understand the mechanics of the golf swing? How will you learn how to adapt your swing, or tell the differences between using a driver and a five iron and a sand wedge? By investigating the underlying mechanics and practicing with those concepts, we hope to gain a better understanding of mathematical induction so that, in the future, we can use it appropriately, identify situations where it would be useful, and, eventually, learn how to *adapt* it to other situations. Of course, it will help to have that domino analogy in mind to guide our intuition, but we should also remember that it is not rigorous mathematics. It also doesn't perfectly describe the other examples we discussed, where a falling domino depended on not only the one immediately behind it, but several others before it.

In the next chapter, we will explore some relevant concepts needed to rigorously state and prove mathematical induction as a proof technique. Specifically, we will study some ideas of *mathematical logic* and investigate how to break down complicated mathematical statements and theorems into their constituent parts, and also how to build interesting and complex statements out of basic building blocks. Along the way, we will introduce some new notation and shorthand that will let us condense some of the wordy statements we make into concise (and precise) mathematical language. With that in hand, we will explore some more fundamental proof strategies, that we will then apply to *everything else we do* in this course, including the induction technique, itself! We will also study some of the ideas of *set theory*, a branch of mathematics that forms the foundation for all other branches. This will be extremely useful for organizing our ideas in the future, but it will also help us define the *natural numbers* in a rigorous manner. With some concepts and knowledge from these two branches of mathematics under our collective belts, we will be able to build mathematical induction on a solid foundation and continue to use it properly.

Chapter 3

Sets: Mathematical Foundations

3.1 Introduction

It's now time for sets education! This might seem like a weird jump to make, after the last chapter. You'll have to trust us when we say that this is actually quite natural and, ultimately, essential. Everything we do in mathematics is built upon the foundation of **sets**, so we better get started talking about them and getting used to them.

3.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to . . .

- Define what a set is, and identify several common examples.
- Use proper notation to define a set and refer to its elements.

- Define and describe several ways to operate on sets; i.e. identify the ways to take two or more sets and create new sets from them.

- Describe how two sets might be compared, as well as apply a proper technique to prove such claims.

- Explain how the natural numbers are related to sets, and relate this to mathematical induction.

3.1.2 Segue from previous chapter

We are building towards a formal statement of mathematical induction as a *theorem*, which we will then prove. To get there, we need to have some fundamental objects to work with and talk about, logically. Sets are those objects! Historically speaking, mathematics was placed on the *set theory* foundation relatively recently, around the turn of the 20th century. Until then, mathematicians tended to "wave their hands" a bit about what was really going on underneath their work; they made a lot of "intuitive" assumptions and hadn't yet tried to rigorously and *axiomatically* describe everything they did. After the work of the mathematician **Georg Cantor** showed everyone some surprising and counter-intuitive results that were perfectly correct and consistent with our assumptions ... well, we realized that we better decide what we've been talking about all along. This is not meant to discredit the work of mathematicians before 1900, of course! We just mean that they were playing a game all along where they hadn't really agreed on a set of rules yet. That's what the **axioms** of set theory are.

3.1.3 Motivation

We are, of course, motivated by our ongoing desire to learn about **proofs**, discover what they are and how they work, and, in particular, rigorize mathematical induction. More generally, though, we are interested in learning about what mathematicians really do, and we are sure that any mathematician in the world can tell you how important **sets** are in their work. They might do so begrudgingly, and say that they themselves could never work in pure *set theory*, but we doubt you'll find anyone who will deny the importance of sets.

Everything we do later on will involve making some claim about a set of objects; that is, we will attempt to say (and subsequently prove) that some fact is True about some particular objects. The way we specify those objects involves **sets**. The way in which we express such facts will involve mathematical logic, and we will get to that soon enough. For now, we need to learn how to express many types of mathematical objects in the first place, before we can even make claims about them.

3.1.4 Goals and Warnings for the Reader

This chapter will likely involve handling some mathematical ideas that are new to you, as opposed to the previous chapters where we focused on puzzles that only relied on some knowledge of numbers and algebra and arithmetic and critical thinking. These new ideas will require careful reading and thinking. As we introduce these concepts and results, we expect you to read through them carefully and do some thinking on the side. Mathematical exposition requires more of the reader than, say, a newspaper article; it expects an *engaged* reader, one who will think carefully about every sentence and sometimes have to pause for a few minutes to ensure full understanding of what has been said so far. Keep this in mind as you read on: reading mathematics can be difficult, but this is to be expected! Don't let it be discouraging; just think of every sentence as a single jigsaw piece in a larger puzzle to be solved.

In particular, don't be surprised if this chapter takes as long (if not longer) to read through (and work through in class) as the previous two chapters combined! The most baffling part of this, as we have observed over the years, is in the **notation** of sets. This is likely the first time in your mathematical careers where you are expected to be as **precise** and **rigorous** as possible with your writing. It is no longer okay to just "have the right idea" in your written work; we really care that you say exactly what you meant to say, and nothing else. As you read what we have written, ask yourself, "Why does that make sense, as opposed to ... something else?" After you have written down an answer to a question or homework problem, read it again and ask yourself, "Does this actually make sense? Does it say what I meant to say, what was in my head? Is someone else guaranteed to read it in the same way I wrote it?"

Also, this chapter will involve some more **abstract** thinking than your typical mathematics course. This might be a shock for you, or maybe not. Either way, this is certainly not material that you can just skim through and expect to pick up on first glance. Now, more than ever, you should take the time and effort to internalize this material. Read some pages and then go think about the material while you eat dinner or shower or play basketball. Try to find examples in real life. Talk with your friends about sets. This may sound silly now but, ultimately, it will help you. Trust us.

3.2 The Idea of a "Set"

A "Collection of Objects" with a Common Property

The intuitive notion of a set is probably not entirely new to you. If you're a baseball card collector, having a "complete set" means owning every single card from a particular printing run by a card manufacturer. If you play board games with friends, you agree on a "set of rules" before playing so there are no unresolved disputes later on. If you performed a laboratory experiment in biology or chemistry or physics class, you collect information into a "data set" and analyze those results to test a hypothesis.

Those are three different situations that each refer to the word **set**, so what is it about that word that relates those contexts and gives it a proper meaning? Essentially, a set refers to a collection of objects of some kind that are grouped together on the basis of having some common property. In the first example, a copy of every baseball card produced by Topps in 1995 would belong to that particular set. In the second example, any agreed-upon convention would belong to your set of rules. In the third example, any data point gathered in your experiment would belong to your data set. In each case, there is a common property that lets us associate particular objects with each other and refer to them as one set.

Sets in Mathematics

Sets are very common, popular, useful and fundamental in mathematics. Because mathematicians work with abstract objects and relationships between those objects, it can be quite difficult to describe exactly what one is thinking about without being able to refer to sets of mathematical objects. We have, in fact, already done so!

For instance, when investigating polynomials and the quadratic formula, we mentioned that a quadratic polynomial $p(x) = ax^2 + bx + c$ with a negative discriminant (when $\frac{b^2}{4a} - c < 0$) will have no roots *in the set of real numbers*. What did we mean? Did you understand that sentence? We were trying to convey the idea that no matter what real number x we choose from among the collection of all real numbers, it would be guaranteed that $p(x) \neq 0$. But what exactly is the set of real numbers? How is it defined? How can we be so sure it even exists? These are actually rather difficult questions to answer, and attempting to do so would take us far off course into the world of set theory.

In the language of mathematics, we aim to be *precise* and *unambiguous* with our sentences and statements, and we seek to establish truths based on certain fundamental assumptions. We need to make those assumptions as a starting point, or else we would have nothing to base our truths upon. These assumptions, that everyone agrees to be part of the "set of rules" before "playing the game" of mathematics, are known as **axioms**.

Perhaps, if you have studied some geometry or read about the Greek mathematician Euclid and his treatise, the *Elements*, then you have heard the word "axiom" before. All of the results of basic geometry that Euclid *proved* were founded upon some basic assumptions: that any two points can be joined by a line segment, that a circle with a given center point and radius must exist, that non-parallel lines intersect, and so on. These statements are simply *agreed upon* to be True at the outset.

Another place we find axioms is in the branch of mathematics known as **set theory**. The axioms of this branch place all of the results involving sets on firm foundations, and using those axioms and results derived from them, we can continue to discover new truths in the mathematical universe. Investigating these axioms and their consequences is better suited for a course devoted to set theory, though, and we will take many of the consequences of the set theoretic

axioms for granted without rigorously proving them. This is not because such proofs are impossible, but merely because they would take too much time and space in this book to accomplish.

What we *will* do is provide a definition of "set" that is satisfactory for the contexts in which we will be using sets in this book. We will also define some basic properties of sets, share some illustrative examples, and dicuss different operations we can perform on sets to create new ones.

3.3 Definition and Examples

3.3.1 Definition of "Set"

Let's start with a definition. As we started to explain above, we often think of sets as being characterized by the objects that are grouped together into that set and the property that makes that grouping make sense. The following definition attempts to make that notion as precise as we possibly can, while also introducing some relevant notation and terminology.

Definition 3.3.1. *A* **set** *is a collection of all objects that have a common, well-defined property. The objects contained in a set are called* **elements** *of the set. The mathematical symbol "$\in$" represents the phrase "is an element of" (and "$\notin$" represents "is not an element of").*

3.3.2 Examples

Let's dive right in with some specific examples of sets (and non-sets, even) to illustrate this definition. It is common in mathematics to use capital letters to refer to *sets* and lowercase letters to refer to *elements* of sets, and we will frequently follow this convention (but not always). To define or describe a set, we need to identify that common, well-defined property that associates the elements of the set with each other. For instance, we could define B to be the set of all teams in Major League Baseball. Is this a well-defined property? If we present you with an object, is ther a definite Yes/No answer to the question of "Does this object have this defining property?" Yes, this is the case here, so this is a property that characterizes a set. (To avoid confusion for readers in the future, let us be more specific and say B refers to MLB teams from the 2012 season.) In the language of mathematics, we would write

$$B = \{\text{Major League Baseball teams from the 2012 season}\}$$

The "curly braces"—{ and }—indicate that the description between them will identify a set, and the text inside is a description of the objects and their common, well-defined property. It now makes sense to say Pittsburgh Pirates $\in$ B and Pittsburgh Penguins $\notin$ B.

Common ways to read the mathematical symbol $\in$ in English are **"is an element of"** or "is a member of" or "belongs to" or "is in". We will mostly

use "is an element of" because it is the least ambiguous of them, and uses the mathematical term **element** appropriately. Any of these other, equivalent, phrases may be used, depending on the context, but are less preferable. (In particular, "is in" can be confused with other set relationships, so we will avoid it entirely, and encourage you to do the same.)

We've also already seen some commonly-used sets of numbers. You know what they are from previous work with these numbers, but you might not usually think of them as sets, which is what they are!

$$\begin{aligned}
\mathbb{N} &= \{\text{ natural numbers }\} = \{1, 2, 3, \ldots\} \\
\mathbb{Z} &= \{\text{ integers }\} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} \\
\mathbb{Q} &= \{\text{ rational numbers }\} \\
&= \{\text{numbers of the form } \frac{a}{b} \text{ , where } a, b \in \mathbb{Z} \text{ and } b \neq 0\} \\
\mathbb{R} &= \{\text{real numbers}\}
\end{aligned}$$

Think about how the second definition of $\mathbb{Q}$ above makes sense. We will see, quite soon, a more condensed way to write out a phrase like "numbers of the form ... blah blah ... with the additional information that ... blah blah". Also, notice that we can't really define $\mathbb{R}$ except to just say they're the real numbers. How do you even define what a real number is? Have you ever tried?

3.3.3 How To Define a Set

Another way of defining or describing a set is simply listing all of its elements. This is convenient when the number of elements in the set is small. For instance, the following definitions of the set V are all *equivalent*:

$$\begin{aligned}
V &= \{\text{A,E,I,O,U}\} \\
V &= \{\text{vowels in the English language}\} \\
V &= \{\text{U,E,I,A,O}\}
\end{aligned}$$

By "equivalent", we mean that each line above defines the *same* set V, using different terms. (Note: we have adopted the convention that y is a consonant, so $y \notin V$.) The common, well-defined property that associated the objects A, E, I, O, and U is the fact that they are all vowels (exhibited in the second definition) and since there are only five such objects, it is simple and convenient to list them all (as in the first definition).

Order and Repetition Don't Matter

Why do you think the third definition is the same as the others? It refers to the same collection of objects, and any set is completely characterized by its elements, so the *order* in which we write the elements *does not matter*. Is U$\in V$? The answer to this question is "Yes", regardless of whether U is written first or last in the list of elements.

Not only does the order of elements not matter within a set, the *repetition* of elements does not matter! That is, the set $A = \{a, a, a\}$ and the set $A = \{a\}$ are exactly the same. Again, remember that a set is completely characterized by its elements; we only care about what is "in" a set. (We will mention this again in Section 3.4.4, when we talk about the "bag analogy" for sets.) Writing $A = \{a, a, a\}$ is just a triply-redundant way of saying $a \in A$ and that *only* a is an element of A. Thus, $A = \{a\}$ is the most concise way of stating the same information.

The Common Property Might Be Being an Element of That Set

Now, still following the idea that we can define a set by writing all of the elements, consider the following definition of a set A:

$$A = \{2, 7, 12, 888\}$$

Surely, this is a set because we just defined it by listing its elements. What is the common, well-defined property that associates its elements, though? With the set V of vowels, we could list the elements *and* provide a linguistic definition, but for this set A, it seems as though we are relegated to listing the elements without knowing a way of *describing* their common property. Mathematically speaking, though, a common property uniting $2, 7, 12, 888$ is that they are all elements of this set A! In the mathematical universe, we have a license for freedom of abstract thought, and merely by discussing this set A and its elements, we have given them that common property. Does this seem satisfactory to you? Can you come up with *another* common, well-defined property that would yield exactly the elements of A? (Hint: identify a polynomial $p(x)$ whose *roots* are exactly $2, 7, 12$, and 888.) If the elements of a set have more than one property that associates them together, do you think it matters which property we have in mind when referring to the set? And what do you think about the set $S := \{2, 7,$ M, Boston Red Sox$\}$? Could there possibly be a common property other than the fact that we have listed them here?

Ellipses Are Sometimes Okay, But Informal

Sometimes, when there is no confusion about the set in question, or it has been defined in another way and we wish to list a few elements as illustrative examples, then it is convenient to use ellipses to condense the listing of elements of a set. For instance, we might write

$$E = \{\text{even natural numbers}\} = \{2, 4, 6, 8, 10, \dots\}$$

This set is *infinitely large*, in fact, so we could not even list all of its elements if we tried, but it is clear enough from the first few elements listed that we are referring to even numbers, especially because we have already referred to E as "the set of even natural numbers". However, we cannot stress enough that this is *not* a precise definition of the set in question. It suffices in an informal context, but it is not mathematically rigorous, and this will become clear as we discuss the following proper way of defining sets.

Set-Builder Notation

The best way to define or describing a set is to identify its elements as particular objects of another set that have a specific property. For instance, if we wished to refer to the set S of all natural numbers between 1 and 100 (inclusive), we could list all of those elements, but this requires a lot of unnecessary writing. We could also use the ellipsis notation $S = \{1, 2, 3, \ldots, 100\}$ but, again, this is not precise without already having a formal definition of S. (Someone might misinterpret the ellipses in a different way.) It is much more precise *and* concise to write

$$S = \{x \in \mathbb{N} \mid 1 \leq x \leq 100\}$$

We read this as "S is the set of all objects x in the set $\mathbb{N}$ of natural numbers *such that* $1 \leq x \leq 100$".

The bar symbol | is read as **"such that"** and indicates that the information to the left tells us what "larger set" the objects come from, while the information to the right tells us the specific property that those objects should have.

(Caution: do *not* use | in other contexts to mean "such that". This is only acceptable in the context of defining sets. It is just used as a place-holder to separate the left side—the set we use to draw elements—from the right side—a description of the property those elements should have.)

This is an example of the very popular and useful **set-builder notation**. We call it this because we are *building* a set by drawing elements from a "larger" set of possibilities, and only including those that have a particular property. To do this, we need to tell the reader (1) what the larger set is, and (2) what the common property is. Let's illustrate this idea with a few examples:

$$\begin{aligned}
S &= \{x \in \mathbb{N} \mid 1 \leq x \leq 100\} = \{1, 2, 3, \ldots, 100\} \\
T &= \{z \in \mathbb{Z} \mid \text{we can find some } k \in \mathbb{Z} \text{ such that } z = 2k\} \\
&= \{\ldots, -4, -2, 0, 2, 4, \ldots\} \\
U &= \left\{x \in \mathbb{R} \mid x^2 - 2 = 0\right\} = \left\{-\sqrt{2}, \sqrt{2}\right\} \\
V &= \left\{x \in \mathbb{N} \mid x^2 - 2 = 0\right\} = \{\ \}
\end{aligned}$$

The last two examples show how the **context** is extremely important. The same common property (satisfying $x^2 - 2 = 0$) can be satisfied by a different set of elements when we change the *larger set* from which we draw elements. Two real numbers satisfy that property, but no natural numbers satisfy it! Do any rational numbers satisfy that property? What do you think?

This is why it is absolutely essential to specify the larger set. A definition like "$U = \left\{x \mid x^2 - 2 = 0\right\}$ is *meaningless* because it is ambiguous and could yield completely different interpretations.

Reading Notation Aloud

We are really learning a new **language** here, and these are some of the basic words and rules of grammar. We'll need some practice translating these sen-

tences into English (in our heads and out loud) and vice-versa. For example, we can say the definition for S above as any of the following, reasonably:

> S is the set of all natural numbers x such that x is between 1 and 100, inclusive.
>
> S is the set of all natural numbers between 1 and 100, inclusive.
>
> S is the set of all natural numbers x that satisfy the inequality $1 \leq x \leq 100$.
>
> S is the set of natural numbers x with the property that $1 \leq x \leq 100$.

Notice that all of them identified the larger set and the common property; the only differences between them are linguistic/grammatical, and they do not alter the mathematical meanings.

Try to write similar sentences for the other definitions. Try getting a verbal definition of a set from a friend and writing down what they said in mathematical symbols.

Consider a definition of the rational numbers $\mathbb{Q}$ that we saw before, and notice that we can rewrite it as:

$$\begin{aligned}\mathbb{Q} &= \left\{ \frac{a}{b} \text{ , where } a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\} \\ &= \left\{ x \in \mathbb{R} \mid \text{we can find } a, b \in \mathbb{Z} \text{ such that } \frac{a}{b} = x \text{ and } b \neq 0 \right\}\end{aligned}$$

Notice the subtle difference between the two definitions. The upper one tells us that all rational numbers are **of the form** $\frac{a}{b}$, and then tells us the particular assumptions of a and b that must be satisfied. The lower one tells us that all rational numbers are *real* numbers with a particular property, namely that we can express that real number as a ratio of integers. We strongly prefer the lower definition, because it tells us more information.

In general, if $P(x)$ represents a sentence (involving English and/or mathematical language) that describes a specific, well-defined *property*, and X is a given set, then the notation

$$S = \{x \in X \mid P(x)\}$$

is read as

> "S is the set of all elements x of the set X such that the property $P(x)$ is true".

In the notation $P(x)$, the letter x represents a variable object, and depending on the particular object we use as x, the property $P(x)$ may hold (i.e. $P(x)$ is true) or may fail (i.e. $P(x)$ is false). If the property holds, then we include x in S (so $x \in S$), and if it fails, we do not include x in S (so $x \notin S$).

Returning to our example of the set E of even natural numbers, it is more precise to write

$$\begin{aligned}E &= \{\text{even natural numbers}\} \\ &= \{x \in \mathbb{N} \mid \text{there is a natural number } n \text{ such that } x = 2n\}\end{aligned}$$

Notice that there are two "layers" of properties here. A natural number is included in our set E if we can find *another* natural number n with the additional property that $x = 2n$. Try to write down a similar definition for the set of odd numbers, or the set of perfect squares. What about the set of primes? The set of palindromic numbers? The set of perfect numbers? Can you write definitions for these sets using set-builder notation?

3.3.4 The Empty Set

What if no elements satisfy a property $P(x)$? What happens then? For instance, consider the definition

$$S = \left\{x \in \mathbb{N} \mid x^2 - 2 = 0\right\}$$

We know that the number x we are "looking for" with that property is $\sqrt{2}$ (and $-\sqrt{2}$, too) but $\sqrt{2} \notin \mathbb{N}$. Thus, no matter what element of $\mathbb{N}$ we let x represent, the property $P(x)$—given by "$x^2 - 2 = 0$"—actually *fails*. Thus, there are no elements of this set. Is this really a set?

Remember, a set is completely characterized by its elements, and a set having *no elements*, such as this one, is characterized by that fact. If we attempted to list its elements, we would end up writing $\{\}$. This set is so special, in fact, that we give it a name and symbol:

Definition 3.3.2. *The* **empty set** *is the set which has no elements. It is denoted by the symbol* $\varnothing$*.*

There are many ways to define the empty set using set-builder notation. (And yes, we do mean *the* empty set; there is only one set with no elements!) We saw one example above, and we're sure you can come up with many others. Consider for example, the following sets:

$$\{a \in \mathbb{N} \mid a < 0\}$$
$$\{r \in \mathbb{R} \mid r^2 < 0\}$$
$$\{q \in \mathbb{Q} \mid q^2 \notin \mathbb{Q}\}$$

Do you see why these all define the same set, the one with no elements?

Context Matters

We should also note again the significance of specifying the larger set X from which we draw our variable element x in a set-builder definition like the one above. For instance, consider the following two sets:

$$S_1 = \{x \in \mathbb{N} \mid |x| = 5\} = \{5\}$$
$$S_2 = \{x \in \mathbb{R} \mid |x| = 5\} = \{-5, 5\}$$

(Note: It is also quite common to **index** sets with the subscript notation you see above, allowing us to use the same letter many times.)

This specification is clearly important, in this case, because it yields two entirely different sets! For this reason, we must be precise and clear when defining a set in this way. A definition like $S = \{x \mid |x| = 5\}$ should be regarded as ambiguous and undesirable, since it leads to issues like the one above.

3.3.5 Russell's Paradox

Perhaps it seems like we are picking nits here, but the reasoning behind our vehemence is rooted in some fundamental ideas of set theory. We wish to avoid some complex issues and paradoxes that might arise without this policy in place. There is a particularly famous example of a paradox involving sets that illustraes why we have the requirement described in the above paragraph, namely that we must specify a larger set when we use set-builder notation. This example is known as *Russell's Paradox* (after the British mathematician Bertrand Russell), and we we will present and discuss it in this section.

Sets Whose Elements Are Sets

First, we should point out that this discussion will introduce the notion that sets can also be *elements* of other sets. This may seem like a strange and far-fetchedly abstract idea right now, but it is a fundamental concept in mathematics.

For a concrete example, think back to our set B of all Major League Baseball Teams. We could also regard each team as a set, where its elements are the players on the team. Thus, it would make sense to say

$$\text{Derek Jeter} \in \text{New York Yankees} \in B$$

since Derek Jeter is an element of the set *New York Yankees*, which is itself an element of the set B. (Notice, however, that Derek Jeter $\notin B$. The relationship signified by "$\in$" is not **transitive**. We will hold off on defining this term until much later. For now, we will point out that the relationship signified by "$\leq$" on the set of real numbers *is* transitive. If we know $x \leq y \leq z$, then we can deduce $x \leq z$. This is *not* the case with the "$\in$" relationship, though.)

Another example is $S = \{1, 2, 3, \{10\}, \varnothing\}$. Yes, the empty set itself can be an element of another set, as can the set $\{10\}$. Why couldn't they? As a thought exercise, we suggest thinking about the difference between $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, and so on. Why are they different sets?

One final example involves the natural numbers $\mathbb{N}$. Let's use $\mathbb{O}$ and $\mathbb{E}$ to represent the *odd* and *even* natural numbers, respectively. What, then, is the set $S = \{\mathbb{O}, \mathbb{E}\}$, and how does it differ, if at all, from $\mathbb{N}$? This is a subtle question, so think about it carefully.

The Paradoxical "Set"

Spend some time on the side thinking about this notion of sets whose elements are sets. For now, though, let us forge ahead with our explanation of Russell's Paradox. Consider the following definition of a "set". We say "set" because it

is actually *not* a properly-defined set, but it remains to be seen exactly why this is the case. When it becomes clear this is not a set, this will be an argument for *requiring* the specification of a larger set to draw from when we use set-builder notation; this is because the definition below does not specify a larger set. Here is that definition:

$$\mathcal{R} = \{x \mid x \notin x\}$$

That's it! Is this a set? What are the elements of $\mathcal{R}$? Think about what the definition above says: the elements of $\mathcal{R}$ are sets that also happen to *not* have themselves as elements. Can you identify any elements of $\mathcal{R}$? Can you identify any objects that are not elements of $\mathcal{R}$?

The first question is far easier to answer: any of the sets we have discussed so far would be elements of $\mathcal{R}$. For instance, the empty set $\varnothing$ contains *no* elements, so it certainly doesn't have *itself* as an element. Thus, $\varnothing \in \mathcal{R}$. Also, notice that$\mathbb{N} \notin \mathbb{N}$ (because the set of natural numbers is not a natural number, itself) and thus $\mathbb{N} \in \mathcal{R}$.

Identifying objects that are *not* elements of $\mathcal{R}$ is a very tricky matter, and we will help you by asking the following question: Is $\mathcal{R}$ an element of itself? Is it true or false that $\mathcal{R} \in \mathcal{R}$? Think about this carefully before reading on. We will walk you through the appropriate considerations.

- Suppose $\mathcal{R} \in \mathcal{R}$ is True.

  The defining property of $\mathcal{R}$ tells us that any of its elements is a set that does *not* have itself as an element. Thus, we can deduce that $\mathcal{R} \notin \mathcal{R}$.

  Wait a minute! Knowing that $\mathcal{R} \in \mathcal{R}$ led us to deduce that, in fact, $\mathcal{R} \notin \mathcal{R}$. Surely, these contradictory facts cannot both hold simultaneously. Accordingly, it must be that our original assumption was bad, so it must be the case that $\mathcal{R} \notin \mathcal{R}$, instead.

- Now, suppose $\mathcal{R} \notin \mathcal{R}$ is True.

  The defining property of $\mathcal{R}$ tells us that any object that is *not* an element of $\mathcal{R}$ must be an element of itself. (Otherwise, it would have been included as an element of $\mathcal{R}$.) Thus, we can deduce that $\mathcal{R} \in \mathcal{R}$.

  Wait a minute! Knowing that $\mathcal{R} \in \mathcal{R}$ led us to deduce that, in fact, $\mathcal{R} \notin \mathcal{R}$. This is also contradictory.

No matter which option we choose—$\mathcal{R} \in \mathcal{R}$ or $\mathcal{R} \notin \mathcal{R}$—we find that the other must also be true, but certainly these contradictory options cannot both be true.

Therein lies the **paradox**. This is not a properly-defined set. If it were, we would find ourselves stuck in the two cases we just saw, and neither of them can be true. It is also not the case that $\mathcal{R}$ is simply $\varnothing$; no, it must be that $\mathcal{R}$ *does not exist* as a set.

The "Set of all Sets" is *Not* a Set

Could we amend the definition of $\mathcal{R}$ somehow to produce the "set" that the definition is trying to convey? What "larger set" should we draw our objects x from so that the definition makes sense and properly identifies a set?

Look back at the English-language interpretation we wrote after the definition: "the elements of $\mathcal{R}$ are sets that also happen to *not* have themselves as elements." The objects x that we wish to test for the desired property ($x \notin x$) are really *all* sets. Perhaps, then, we should just define X to be the set of all sets, and use the phrase "$x \in X$" as part of our definition of $\mathcal{R}$. That would fix it, right?

$$\mathcal{R} = \{x \in X \mid x \notin x\}$$

Well, no, not at all! **The "set of all sets" is, itself, *not* a set**. If it were, this would lead us to exactly the same paradox as before! Nothing would be different, except we would have explicitly stated the "larger set" from which we draw objects x that was previously left *implicitly*-specified.

The main issue is that not specifying a "larger set" from which to draw objects, or implicitly referring to the "set of all sets", results in this type of undesirable paradox. Accordingly, we must not allow such definitions. Any attempt to define a set that draws objects x from the "set of all sets", whether implicitly or explicitly, is not a *proper definition* of a set.

Further Discussion

There is nothing inherently wrong with the property $P(x)$ given by "$x \notin x$", though. The issue is with that "larger set" we use. For instance, take the set

$$\mathcal{S} = \left\{ x \in \left\{ \frac{1}{2}, \frac{3}{4}, \frac{5}{2} \right\} \,\middle|\, x \notin x \right\}$$

What are its elements? The only possibilities are those elements drawn from the larger set $\left\{\frac{1}{2}, \frac{3}{4}, \frac{5}{2}\right\}$. Notice that none of those numbers are sets that contain themselves as elements. Thus, this is a proper definition of the set $\left\{\frac{1}{2}, \frac{3}{4}, \frac{5}{2}\right\}$, itself! With the previous definition for $\mathcal{R}$, the object we were attempting to define was allowed as one of the variable objects x in its own definition, and that is where the issue arose.

We hope that we haven't led your thoughts too far astray from the original discussion of examples of sets, but we felt it was important to point out that it is possible to construct ill-defined "collections" that are not sets in the mathematical sense of the word. For the most part, we will not face any such issues with the sets we work with in this book, but to gloss over these issues or simply not mention them at all would be unfair to you, as a student. If you find yourself interested in these issues, seek out an introductory book on set theory.

There are other ways that definitions of "sets" can be ill-formed, as well, but the ensuing examples are based on (English) linguistic issues and not any mathematical underpinnings, as in Russell's Paradox. For instance, we could

say "Let N be the set of all classic novels from the 20th century." Being a "classic novel" is *not* a well-defined property, and cannot be used to identify elements of such a set. The notion of a "classic" is subjective and not rigorously precise. Also, we could say "Let B be the set of people who will be born tomorrow" but this temporal dependence in the definition ensures that we will never actually know what the elements of B are. When tomorrow arrives, the definition will then refer to the next day, and so on. Can you come up with some other examples of ill-formed "collections" of elements? Can you come up with any paradoxes like the one above?

In general, the following statement is the most important idea to take away from this discussion of Russel's Paradox:

> Under the agreed-upon rules of sets (the axioms of set theory), there is **no** set of all sets.

3.3.6 Standard Sets and Their Notation

We have referred to and used some common sets of numbers already, so we will list now some sets and their standard symbols:

- The *natural numbers*: $\mathbb{N} = \{1, 2, 3, 4, \dots\}$
- The first n natural numbers: $[n] := \{1, 2, 3, \dots, n-1, n\}$
- The *integers*: $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- The *rational numbers*: $\mathbb{Q} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z} \text{ and } n \neq 0 \right\}$
- The *real numbers*: $\mathbb{R}$
- The *complex numbers*: $\mathbb{C}$

We have used $\mathbb{N}$ and $\mathbb{Z}$ a few times already. The rational numbers $\mathbb{Q}$ (we use $\mathbb{Q}$ since $\mathbb{R}$ was already taken, and rational numbers are *quotients*) are all of the *fractions*, or ratios of integers, both positive and negative. The real numbers are harder to describe. Why could we not list a few elements, like we did with $\mathbb{N}$ and $\mathbb{Z}$? Why is it that $\mathbb{R} \neq \mathbb{Q}$? For now, we essentially take for granted our collective knowledge of these sets of numbers, but think for a minute about that. (We mention the complex numbers $\mathbb{C}$ because you might be familiar with them, but we will not have occasion to use them in this book.)

How do we *know* that a set like $\mathbb{N}$ exists? Why is it that we think of $\mathbb{R}$ as a number line? How many "more" elements are there in $\mathbb{Z}$, as compared to $\mathbb{N}$? How many "more" elements are there in $\mathbb{R}$, as compared to $\mathbb{Q}$? Can we even answer these questions? In the very near future, we will rigorously derive the set $\mathbb{N}$ and prove that it exists as the only set with a particular property. This will be essential when we return to our investigation of mathematical induction. (Remember our goals from that chapter?)

3.3.7 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What does the symbol "$\in$" mean?

(2) How would you read the statement "$x \in S$" out loud?

(3) Is it possible for a set to be an element of another set? If so, give an example. Is it possible for a set to be an element of itself?

(4) How would read the statement "$\{x \in \mathbb{N} \mid x \leq 5\}$" out loud? Can you list the elements of this set?

(5) What is this set: $\{z \in \mathbb{Z} \mid z \in \mathbb{N}\}$?

(6) What is this set: $\{x \in [10] \mid x \geq 7\}$?

(7) For each of the following sets, state *how many* elements they have:

   (a) $\varnothing$

   (b) $\{1, 2, 10\}$

   (c) $\{1, \varnothing\}$

   (d) $\{\, \varnothing \,\}$

(8) Is $x \in \{\, 1, 2, \{x\} \,\}$? Is $\{x\} \in \{\, 1, 2, \{x\} \,\}$?

(9) Let $A = \{a, b, c\}$ and $B = \{b, c, a\}$ and $C = \{a, a, b, c, a, b\}$. Are these sets equal or not?

(10) Is $\mathbb{Z} = \mathbb{Q}$? Why or why not?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Write a definition of the set of natural numbers that are multiples of 4 using set-builder notation.

(2) Consider the set $S = \{3, 4, 5, 6\}$. Define S in two different ways using set-builder notation.

(3) Give an example of a set X that satisfies $\mathbb{N} \in X$ but $\mathbb{Z} \notin X$.

(4) Give an example of a set with 100 elements.

(5) Give an example of a sets A, B, C such that $A \in B$ and $B \in C$ but $A \notin C$.

(6) Write a definition of the set of *odd integers* using set-builder notation.

(7) Write a definition of the set of integers that are not natural numbers using set-builder notation.

(8) Consider the following sets:

$$A = \left\{x \in \mathbb{R} \mid x^2 - 3x + 2 \geq 0\right\}$$
$$B = \{y \in \mathbb{R} \mid y \leq 1 \text{ or } y \geq 2\}$$

Explain why $A = B$.

(9) Consider the following sets:

$$C = \left\{x \in \mathbb{R} \mid x^2 - 4 \geq 0\right\}$$
$$D = \{y \in \mathbb{R} \mid y \geq 2\}$$

Is $C = D$? Why or why not? Write your explanation with good mathematical notation, using $\in$ and $\notin$.

(10) Try explaining Russell's Paradox to a friend, even one who does not study mathematics. What do they understand about it? Do they object? Do their ideas make sense? Have a discussion!

3.4 Subsets

3.4.1 Definition and Examples

Let's discuss a topic whose basic idea we have already been using. Specifically, let's investigate the notion of **subsets**.

Definition 3.4.1. *Given two sets A and B, if every element of A is also an element of B, then we say A is a* **subset** *of B.*

The mathematical symbol for subset is $\subseteq$, so we would write $A \subseteq B$.

If we want to indicate that A is a subset of B but is also not equal *to B, we would write $A \subset B$ and say that A is a* **proper subset** *of B.*

We can also write these relationships as $B \supseteq A$, or $B \supset A$, respectively. In these cases, we would say B is a **superset** *of A or B is a* **proper superset** *of A, respectively.*

Notice the similarities between these symbols and the *inequality* symbols we use to compare real numbers. We write inequalities like $x \leq 2$ or $5 > z > 0$ and understand what those mean based on the "direction" of the symbol and whether we put a bar underneath it. The symbols $\subseteq, \subset, \supseteq, \supset$ work exactly the same way, except they refer to "containment of elements" as opposed to "magnitude of a number".

Standard Sets of Numbers

The standard sets of numbers we mentioned in the last section are related via the subset relation quite nicely. Specifically, we can say

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

Again, we take for granted our collective knowledge of these sets of numbers to allow us to make these claims. However, there are some profound and intricate mathematical concepts involved in describing exactly why, say, the set $\mathbb{R}$ exists and is a proper superset of $\mathbb{Q}$. For now, though, we use these sets to illustrate the **subset** relationship.

Since we know the subset relationships above are **proper**, we used that corresponding symbol, "$\subset$". In general, it is common in mathematical writing to simply use the "$\subseteq$" symbol, even if it is known that "$\subset$" would apply. We might only resort to using the "$\subset$" symbol when it is important, in context, to indicate that the two sets are *not equal*. If that information is not essential to the context at hand, then we might just use the "$\subseteq$" symbol.

Set-Builder Notation Creates Subsets

One way that we have already "used" the idea of a subset was in our use of set-builder notation. This is used to define a set to be all of the elements of a "larger" set that satisfy a certain property. We define a property $P(x)$, draw a variable object x from a larger set X, and include any elements x that satisfy the property $P(x)$. Notice that any element of this new set must be an element of X, simply based on the way we defined it. Thus, the following relationship holds

$$\{x \in X \mid P(x)\} \subseteq X$$

regardless of the set X and the property $P(x)$. Depending on the set X and the property $P(x)$, it may be that the proper subset symbol $\subset$ applies but, in general, we can say for sure that $\subseteq$ applies.

Try to come up with some examples of sets X and properties $P(x)$ so that $\subseteq$ applies, then try to come up with some examples where $\subset$ applies. Try to come up with one set X and two different properties $P_1(x)$ and $P_2(x)$ so that $\subset$ applies for $P_1(x)$ and $\subseteq$ applies for $P_2(x)$. Try to identify two *different* sets X_1 and X_2 and two *different* properties $P_1(x)$ and $P_2(x)$ so that

$$\{x \in X_1 \mid P_1(x)\} = \{x \in X_2 \mid P_2(x)\}$$

Can you do it?

Examples

A set is a subset of another set if and only if every element of the first one is an element of the second one. For instance, this means that the following claims all hold:

$$\{142, 857\} \subseteq \mathbb{N}$$
$$\left\{\sqrt{3}, -\pi, 8.2\right\} \subseteq \mathbb{R}$$
$$\left\{x \in \mathbb{R} \mid x^2 = 1\right\} \subseteq \mathbb{Z}$$

Do you see why these are True?

For a subset relationship to fail, then, we must be able to find an element of the first set that is *not* an element of the second set. For instance, this means that the following claims all hold:

$$\{142, -857\} \not\subseteq \mathbb{N}$$
$$\left\{\sqrt{3}, -\pi, 8.2\right\} \not\subseteq \mathbb{Q}$$
$$\left\{x \in \mathbb{R} \mid x^2 = 5\right\} \not\subseteq \mathbb{Z}$$

Finding All Subsets of a Set

Let's work with a specific set for a little while. Define $A = \{1, 2, 3\}$. Can we identify *all* of the subsets of A? Sure, why not?

$$\{1\} \subseteq A$$
$$\{2\} \subseteq A$$
$$\{3\} \subseteq A$$
$$\{1, 2\} \subseteq A$$
$$\{1, 3\} \subseteq A$$
$$\{2, 3\} \subseteq A$$
$$A = \{1, 2, 3\} \subseteq A$$
$$\varnothing \subseteq A$$

Identifying the first six subsets was fairly straightforward, but it's important to remember that A and $\varnothing$ are subsets, as well. (Notice: it is true, in general, that for any set S, $S \subseteq S$ and $\varnothing \subseteq S$. Think about this!)

Consider the set B whose elements are all of the sets we listed above:

$$B = \left\{ \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, A, \varnothing \right\}$$

It is true that any element $X \in B$ satisfies $X \subseteq A$. Do you see why?

3.4.2 The Power Set

This process of identifying all subsets of a given set is common and useful, so we identify the resulting set with a special name.

Definition 3.4.2. *Given a set A, the* **power set** *of A is defined to be the set whose elements are all of the subsets of A. It is denoted by $\mathcal{P}(A)$.*

Our parenthetical observation from the above paragraph tells us that $S \in \mathcal{P}(S)$ and $\varnothing \in \mathcal{P}(S)$, for any set S.

Look back at our example set $A = \{1, 2, 3\}$ above. What do you notice about the number of elements in $\mathcal{P}(A)$? How does it relate to the number of elements in A? Do you think there is a general relationship between the number of elements in S and $\mathcal{P}(S)$, for an arbitrary set S?

Example 3.4.3. Let's find $\mathcal{P}(\varnothing)$. What are the subsets of the empty set? There is only one, the empty set itself! (That is, $\varnothing \subseteq \varnothing$, but no other set satisfies this.) Accordingly, the power set $\mathcal{P}(\varnothing)$ has one element, the empty set:

$$\mathcal{P}(\varnothing) = \{\, \varnothing \,\}$$

Notice that this is *different* from the empty set itself:

$$\varnothing \neq \{\, \varnothing \,\}$$

Why is this true? Compare the elements! The empty set has *no* elements, but the set on the right has exactly *one* element. (In general, this can be a helpful way to compare two sets.) To give you some practice, we'll point out that would read the above line aloud as:

> "The empty set and the set containing the empty set are two different sets."

Example 3.4.4. Let's try this process with another set, say $A = \{\, \varnothing, \{1, \varnothing\} \,\}$. We can list the elements of $\mathcal{P}(A)$ as

$$\mathcal{P}(A) = \{\, \{\varnothing\}, \{\{1, \varnothing\}\}, \{\varnothing, \{1, \varnothing\}\}, \varnothing, \,\}$$

This may look strange, with all of the empty sets and curly braces, but it is important to keep the subset relationships straight. It is true, in this example, that

$$\varnothing \in A, \quad \{\varnothing\} \subseteq A, \quad \{\varnothing\} \in \mathcal{P}(A), \quad \{\varnothing\} \subseteq \mathcal{P}(A)$$

Why are these relationships true? Think carefully about them, and try to write a few more on your own. The distinction between "$\in$" and "$\subseteq$" is very important!

3.4.3 Set Equality

When are two sets equal? The main idea is that two sets are equal if they contain "the same elements", but this is not a precise definition of equality. How can we describe that property more explicitly and rigorously? To say that two sets, A and B, have "the same elements" means that any element of A is also an element of B, and every element of B is also an element of A. If both of these properties hold, then we can be guaranteed that the two sets contain precisely the same elements and are, thus, equal. If you think about it, you'll notice that we can phrase this in terms of **subsets**. How convenient!

Definition 3.4.5. *We say two sets, A and B, are* **equal***, and write $A = B$, if and only if $A \subseteq B$ and $B \subseteq A$.*

(What happens if we use the $\subset$ symbol in the definition, instead of $\subseteq$? Is this the same notion of set equality? Why or why not?)

This definition will be very useful in the future when we learn how to *prove* two sets are equal and we can't simply list the elements of each and compare them. By constructing two arguments and proving the subset relationship in "both directions", we can show that two sets are equal. For now, let's see this definition applied to a straightforward example.

Example 3.4.6. How can we use this to definition to observe that the following equality holds?

$$\{x \in \mathbb{Z} \mid x \geq 1\} = \mathbb{N}$$

We just need to see that the $\subseteq$ and $\supseteq$ relationship applies between the two sides. First, is it true that every integer that is at least 1 is a natural number? Yes! This explains why

$$\{x \in \mathbb{Z} \mid x \geq 1\} \subseteq \mathbb{N}$$

Second, is it true that every natural number is a positive integer that is at least 1? Yes! This explains why

$$\{x \in \mathbb{Z} \mid x \geq 1\} \supseteq \mathbb{N}$$

Together, this shows that the equality stated above is correct.

3.4.4 The "Bag" Analogy

It has been our experience that **sets** are a rather difficult notion to grasp when they are introduced. Specifically, the **notation** associated with sets throws students for a loop, and they end up writing down things that make no sense! It is essential to keep straight the differences between the symbols $\in$ and $\subseteq$.

Here is a helpful analogy to keep in mind: a set is like a **bag** with some stuff in it. The bag itself is irrelevant; we just care about what *kind* of stuff inside (i.e. what the elements are). Think of the bag as a non-descript plastic bag you'd get at the grocery store, even. All those bags are identical; to distinguish between any two bags, we need to know what kind of things are *inside* them.

If I put an apple and an orange in a grocery bag, surely it doesn't matter in what *order* I put them in. All you would need to know is that I have some apples and oranges. It also doesn't matter how *many* apples or oranges I have in the bag, because we only care about what *kind* of stuff is in there. Think of it as answering questions of the form "Are there any _____s in the bag? Yes or no?" It doesn't matter if I have two identical apples or seven or just one in my bag; if you ask me whether I have *any* apples, I'll just say "Yes". This is related to the notion that the order and repetition of elements in a set don't matter. A set is completely characterized by what its elements are.

This also helps when we think about sets as elements of other sets, themselves. Who's to stop me from just putting a whole bag inside another bag¿ Look at the set A we defined in the example above:

$$A = \{ \varnothing , \{1, \varnothing\} \}$$

This set A is a bag. What's inside the bag? There are two objects inside the bag (i.e. there are two elements of A). They both happen to be bags, themselves! One of them is a plain-old empty bag, with nothing inside it. (That's the empty set.) Okay, that's cool. The other one has two objects inside it. One of those objects is the number 1. Cool. The other such object is another empty bag.

Distinguishing "$\in$" and "$\subseteq$"

This analogy also helps with understanding the differences between "$\in$" and "$\subseteq$". Keep the example A in mind again. When we write $x \in A$, we mean x is an object inside the bag A. If we peeked into A, we would see an x sitting there at the bottom amongst the stuff. Let's use this idea to compare two examples.

- We see that $\varnothing \in A$ is true here. If we look inside the bag A, we see an empty bag amongst the stuff (the elements).
- We also see that $\{\varnothing\} \notin A$ is true here. If we look inside the bag A, we don't see a bag containing *only* an empty bag. (This is what $\{\varnothing\}$ is, mind you: an empty bag inside another bag.)

  Do you see such an object? Where? I defy you to show me, amongst the stuff inside the bag A, a bag containing *only* an empty bag.

  What do I see inside the bag A? Well, I see two things: an empty bag, and a bag that has *two* objects inside it (an empty bag, and the number 1). Neither of those objects is what we were looking for!

When we write $X \subseteq A$, we mean that the two bags, X and A, are somehow comparable. Specifically, we are saying that all of the stuff inside X is also stuff inside A. We are really rooting through all of the objects inside X, taking them out one by one, and making sure we also see that object inside A. Let's use this idea to compare two examples.

- We see that $\{\varnothing\} \subseteq A$ is true here. We are *comparing* the bag on the left with the bag on the right. What stuff is inside the bag on the left?

There's just one object in there, and it is an empty bag, itself. Now, we peek inside A. Do we also see an empty bag in there? Yes we do! Thus, the "$\subseteq$" symbol applies here.

- We also see that $\{1\} \not\subseteq A$ is true here. To compare these two bags, we'll pull out an object from the bag on the left and see if it is also in the bag A. Here, we only have one object to pull out: the number 1. Now, let's peek inside the bag A. Do we see a 1 sitting in there amongst the stuff? No, we don't!

  We would have to peek *inside* something at the bottom of the bag A to find the number 1; that number isn't just sitting out in plain sight. Thus, $\{1\} \not\subseteq A$.

Look back over some examples we have discussed already with this new analogy in mind. Does it help you understand the definitions and examples? Does it help you understand the difference between "$\in$" and "$\subseteq$" and "$\supseteq$"? If not, can you think of another analogy that would help you?

3.4.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) Is $\mathbb{N} \subseteq \mathbb{R}$? Is $\mathbb{R} \subseteq \mathbb{N}$? Is $\mathbb{Q} \subseteq \mathbb{Z}$? Why or why not?

(2) What is the difference between $\subset$ and $\subseteq$? Give an example of sets A, B such that $A \subseteq B$ is True but $A \subset B$ is False.

(3) What is the difference between $\in$ and $\subseteq$? Give an example of sets C, D such that $C \subseteq D$ but $C \notin D$.

(4) Let S be any set. What is the **power set** of S? What type of mathematical object is it? How is it defined?

(5) Suppose $S \subseteq T$. Does this mean $S = T$? Why or why not?

(6) Explain why $\varnothing \subseteq S$ and $\varnothing \in \mathcal{P}(S)$ for any set S.

(7) Suppose $X \in \mathcal{P}(A)$. How are X and A related, then?

(8) Is it possible for $A = \mathcal{P}(A)$ to be true? (This one is rather tricky, but think about it!)

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Write out the elements of the set $\mathcal{P}(\mathcal{P}(\varnothing))$.

(2) Write out the elements of the sets $\mathcal{P}([1])$ and $\mathcal{P}([2])$ and $\mathcal{P}([3])$. Can you make a conjecture about how many elements $\mathcal{P}([n])$ has? (Can you prove it? We don't expect you to *now*, but soon enough; think about it!)

(3) Let $A = \{x, \heartsuit, \{4\}, \varnothing\}$. For each of the following statements, decide whether it is True or False and briefly explain why.

(a) $x \in A$
(b) $x \subseteq A$
(c) $\{x, \heartsuit\} \subseteq A$
(d) $\{x, \varnothing\} \subset A$
(e) $\{x, \heartsuit, z, 7\} \supseteq A$
(f) $\{x\} \in \mathcal{P}(A)$
(g) $\{x\} \subseteq \mathcal{P}(A)$
(h) $\{\heartsuit, x\} \in \mathcal{P}(A)$
(i) $\{4\} \in \mathcal{P}(A)$
(j) $\{\varnothing\} \in \mathcal{P}(A)$
(k) $\{\varnothing\} \subseteq \mathcal{P}(A)$

Hint: 7 of these are True, and 4 are False.

(4) Give an example of sets A, B such that $A \in B$ and $A \subseteq B$ are both true.

(5) Is $\{1, 2, 12\} \subseteq \mathbb{R}$?

(6) Is $\{-5, 8, 12\} \subseteq \mathbb{N}$?

(7) Is $\{1, 3, 7\} \in \mathcal{P}(\mathbb{N})$?

(8) Is $\mathbb{N} \in \mathcal{P}(\mathbb{Z})$?

(9) Is $\mathcal{P}(\mathbb{N}) \subseteq \mathcal{P}(\mathbb{Z})$? Are they equal sets? Why or why not?

(10) Give an example of an infinite set T such that $T \in \mathcal{P}(\mathbb{Z})$ but $T \notin \mathcal{P}(\mathbb{N})$.

(11) Suppose G, H are sets and they satisfy $\mathcal{P}(G) = \mathcal{P}(H)$. Can we conclude that $G = H$? Why or why not? (Don't try to formally prove this; just think about it and try to talk it out.)

(12) Give an example of a set W such that $W \subseteq \mathcal{P}(\mathbb{N})$ but $W \notin \mathcal{P}(\mathbb{N})$.

3.5 Set Operations

When you first learned about numbers, a natural next step was to learn about how to *combine* them: multiplication, addition, and so on. Thus, a natural next step for us now is to investigate how we can take two sets and *operate* on them to produce other sets. How can we *combine* sets in interesting ways? There are several such operations that have standard, notational symbols and we will introduce you to those operations now.

Throughout this section, we assume that we are given two sets A and B that are each subsets of a larger **universal set** U. That is, we assume $A \subseteq U$ and $B \subseteq U$. The reason we make this assumption is that each operation involves defining another set by identifying elements of a larger set with a specific property, so we must have some set U that is guaranteed to contain all of the elements of A and B so we can even work with those elements. (Again, ensuring this may seem nit-picky, but it is meant to avoid nasty paradoxes like the one we investigated before.) Assuming those sets A, B, U exist, though, we can forge ahead with our definitions.

3.5.1 Intersection

This operation collects the elements common to two sets and includes them in a new set, called the **intersection**.

Definition 3.5.1. *Let A and B be any sets. The* **intersection** *of A and B is the set of elements that belong to both A* and *B, and is denoted by $A \cap B$. Symbolically, we define*

$$A \cap B = \{x \in U \mid x \in A \text{ and } x \in B\}$$

Example 3.5.2. Define the following sets:

$$\begin{aligned} S_1 &= \{1, 2, 3, 4, 5\} \\ S_2 &= \{1, 3, 7\} \\ S_3 &= \{2, 4, 7\} \\ U &= \mathbb{N} \end{aligned}$$

Then, we see that, for example,

$$\begin{aligned} S_1 \cap S_2 &= \{1, 3\} \\ S_1 \cap S_3 &= \{2, 4\} \\ S_2 \cap S_3 &= \{7\} \end{aligned}$$

Also, since $S_1 \cap S_2$ is, itself, a set, it makes sense to consider $(S_1 \cap S_2) \cap S_3$. However, those two sets share no elements, so we write

$$(S_1 \cap S_2) \cap S_3 = \varnothing$$

The situation where two sets have no common elements, as seen in the above example, is common enough that we have a specific term to describe such sets:

Definition 3.5.3. *If $A \cap B = \varnothing$, then we say A and B are* **disjoint**.

Intersections and Subsets

You might have observed already that we can say $A \cap B \subseteq A$ and $A \cap B \subseteq B$, no matter what A and B are. Let's prove this fact!

Proposition 3.5.4. *Let A and B be any sets. Then,*

$$A \cap B \subseteq A$$

and

$$A \cap B \subseteq B$$

By the way, a **proposition** like this is just a "mini result". It's not difficult or important enough to be called a theorem, but it does require a little proof.

Proof. Let's say we have two sets, A and B. To prove a subset relationship, like $A \cap B \subseteq A$, we need to show that every **element** of the set on the left ($A \cap B$) is also an element of the set on the right (A).

Let's consider an arbitrary element $x \in A \cap B$. By the definition of $A \cap B$, we know that $x \in A$ and $x \in B$. Thus, we know that $x \in A$. This was our goal, so we have shown that $A \cap B \subseteq A$.

Also, we know that $x \in B$, so we have also shown that $A \cap B \subseteq B$. □

This might seem like a simple observation and an easy proof, but we still need to go through these logical steps to rigorously explain why those subset relationships hold true. Also, notice the type of **proof structure** we used here. To prove a subset relationship holds true, we need to consider an **arbitrary element** of one set and deduce that it is also an element of the other set. This will be our method for proving any claim about subsets.

What if $A \subseteq B$? What can we say about $A \cap B$, in relation to A and B? Try to prove a statement about this!

3.5.2 Union

This operation collects the elements of either of two sets and includes them in a new set, called the **union**.

Definition 3.5.5. *Let A and B be any sets. The* **union** *of A and B is the set of elements that belong to either A* or *B, and is denoted by $A \cup B$. Symbolically, we define*

$$A \cup B = \{x \in U \mid x \in A \text{ or } x \in B\}$$

Notice that the "or" in the definition is an *inclusive* "or", meaning that $A \cup B$ includes any element that belongs to A or B or possibly both sets.

Example 3.5.6. Returning to the sets S_1, S_2, S_3 we defined above in Example 3.5.2, we can say

$$S_1 \cup S_2 = \{1, 2, 3, 4, 5, 7\}$$
$$S_1 \cup S_3 = \{1, 2, 3, 4, 5, 7\}$$
$$S_2 \cup S_3 = \{1, 2, 3, 4, 7\}$$

Also, since each of these unions are sets themselves, we could find their union with another set. For example,

$$(S_1 \cup S_2) \cup S_3 = \{1, 2, 3, 4, 5, 7\} \cup \{2, 4, 7\} = \{1, 2, 3, 4, 5, 7\}$$

Unions and Subsets

Notice that $A \subseteq (A \cup B)$ and $B \subseteq (A \cup B)$, no matter what A and B are. Let's prove that!

Proposition 3.5.7. *Let A and B be any sets. Then,*

$$A \subseteq (A \cup B)$$

and

$$B \subseteq (A \cup B)$$

Proof. Let's say we have two sets, A and B. To prove $A \subseteq (A \cup B)$, we need to show that every element of A is also an element of $A \cup B$.

Let $x \in A$ be arbitrary and fixed. Then it is certainly that $x \in A$ or $x \in B$ (since $x \in A$). This shows $x \in A \cup B$. Since x was arbitrary, we have shown $A \subseteq A \cup B$.

Let $y \in B$ be arbitrary and fixed. Then it is certainly true that $y \in A$ or $y \in B$ (since we already know $y \in B$). This shows $y \in A \cup B$. Since y was arbitrary, we have shown $B \subseteq A \cup B$. □

What can you say about the relationship between $A \cap B$ and $A \cup B$? If $A \subseteq B$, can we say anything about the relationship between B and $A \cup B$? Try to prove your observations!

We should emphasize that claims like this—that $A \subseteq A \cup B$ for any sets A and B—need to be proven; they do not hold **by definition**. The definition of the union of two sets is given above. Notice it says nothing about how A and $A \cup B$ are related; it just tells us what the object $A \cup B$ actually is. When you invoke or cite a definition and use it, be sure to do so; but also, be sure to explain any claim that isn't exactly a definition. Since we have proven these two little lemmas, we get to use them in the future by referencing them; had we not done so, we would have to re-explain these little facts every time we try to invoke them!

3.5.3 Difference

This operation takes the elements of one set and removes the elements that also belong to another set.

Definition 3.5.8. *The* **difference** *between A and B, denoted by $A - B$, is the set of all elements of A that are* not *elements of B. Symbolically, we define*

$$A - B := \{x \in U \mid x \in A \text{ and } x \notin B\}$$

Example 3.5.9. Returning to the sets S_1, S_2, S_3 we defined above in Example 3.5.2, we can say, for example, that

$$S_1 - S_2 = \{2, 4, 5\}$$
$$S_2 - S_1 = \{7\}$$
$$S_2 - S_3 = \{1, 3\}$$

Set Difference Is *Not* Symmetric

Notice that $S_1 - S_2 \neq S_2 - S_1$ in the example above. In general, the operation "$-$", in the context of sets, is not **symmetric**, and this example here shows that. Can you find two sets A, B so that $A - B = B - A$? Can you find two sets A, B so that $A - B = B - A \neq \varnothing$?

Each of the other operations we have defined thus far is, in fact, symmetric. That is, $A \cap B = B \cap A$ and $A \cup B = B \cup A$. Look back at the definitions for these operations and see why this makes sense. What is it about the *language* used in the property definition of the operation that makes this true?

Notation Notes

One more comment on this set difference notation. Notice that we use the standard subtraction symbol, "$-$", but this has nothing to do with "subtraction" in the way we usually think of it, like with numbers. This might be the first time you've encountered this ambiguity, or perhaps not, but there is a larger point that is relevant to mathematical notation and terminology: many symbols have different meanings depending on the *context*.

When we write $7 - 5$, we clearly mean subtraction, i.e. $7 - 5 = 2$. However, when we write $A - A$ where A has been identified as a *set*, we mean the set difference operation, i.e. $A - A = \varnothing$. Be sure to check the context of any statement to ensure that the symbols therein do mean what you think they mean!

3.5.4 Complement

This operation identifies all of the elements that lie "outside" a set. This operation depends on the context of the universal set U. You'll notice that this is evident in the deifnition, and we will illustrate this through examples, as well.

Definition 3.5.10. *The* **complement** *of A is the set of all elements that are* not *elements of A, and is denoted by $\bar{A}$. Symbolically, we define*

$$\bar{A} = \{x \in U \mid x \notin A\}$$

Remember that we assumed A, B, U are given sets that satisfy $A \subseteq U$ and $B \subseteq U$. Within this context, the set $\bar{A}$ is well-defined, but this set certainly depends on A and U!

Example 3.5.11. For instance, let's return to the sets S_1, S_2, S_3 we defined above in Example 3.5.2. There, we used the context $U = \mathbb{Z}$. In that case,

$$\bar{S}_1 = \{6, 7, 8, 9, \dots\}$$

However, what if we had used $U = \{1, 2, 3, 4, 5, 6, 7\}$? In that case,

$$\bar{S}_1 = \{6, 7\}$$

Since the symbolic notation $\bar{A}$ makes no indication of the set U that it depends on, it is important to make this set clear in whatever context we are working. Try identifying some sets A, U_1, U_2 so that $\bar{A}$ in U_1 is different from $\bar{A}$ in U_2, and try identifying some sets so that $\bar{A}$ is the same in both cases.

3.5.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between a union and intersection of two sets?

(2) What does it mean for two sets to be disjoint?

(3) What is $\mathbb{Z} \cap \mathbb{N}$? What is $\mathbb{Z} \cup \mathbb{N}$? What is $\mathbb{Z} - \mathbb{N}$?

(4) Is it possible for $A - B = B - A$ to be true? How?

(5) What is $\overline{[3]}$ in the context of $\mathbb{N}$? What about in the context of $\mathbb{Z}$? Of $\mathbb{R}$? Try writing your answers using good mathematical notation, and set-buidler notation, perhaps.

(6) Is $(A \cap B) \cap C = (A \cap B) \cap C$ always true? Why or why not? What about with $\cup$ instead of $\cap$?

(7) What is the difference between the statements "$7 - 5$" and "$[7] - [5]$"?

(8) Suppose $x \in A$. Does $A - x$ make sense, notationally? How can you change it to make sense?

(9) What is $(\mathbb{Z} - \mathbb{N}) \cup \mathbb{R}$?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) List the elements of the following sets:

(a) $[7] \cup [10]$

(b) $[10] \cap [7]$

(c) $[10] - [7]$

(d) $([12] - [3]) \cap [8]$

(e) $(\mathbb{N} - [3]) \cap [7]$

(f) $(\mathbb{Z} - \mathbb{N}) \cap \mathbb{N}$

(g) $\overline{\mathbb{N}} \cap \{0\}$, in the context of $\mathbb{Z}$

(2) Find an example of sets A, B, C such that $(A - B) - C = A - (B - C)$. Then, find an example where they are **not** equal.

(3) State and prove a relationship between $\overline{A}$ and $U - A$.

(4) Let $A = [12]$, let E be the set of even integers, and let P be the set of prime natural numbers. What is $A \cap E$? What is $A \cap P$? What is $(A \cap E) \cap P$? Is it the same as $A \cap (E \cap P)$?

Suppose the context is $U = \mathbb{N}$. What are $\overline{A \cap E}$ and $\overline{A} \cap \overline{E}$?

(5) What is $\{1\} \cap \mathcal{P}(\{1\})$?

(6) Consider the sets $\{1\}$ and $\{2, 3\}$. Compare the sets $\mathcal{P}(\{1\} \cup \{2, 3\})$ and $\mathcal{P}(\{1\}) \cup \mathcal{P}(\{2, 3\})$. What do you notice?

Repeat this exercise with "$\cap$" instead of "$\cup$". What do you notice?

(7) Let A, U be sets, and suppose $A \subseteq U$. Let $B = \overline{A}$, in the context of U. What do you think $\overline{B}$ is? Why?

3.6 Indexed Sets

3.6.1 Motivation

Let's discuss a notion that we referenced briefly before and have been using already: the concept of **indexing** a collection of sets. This type of notation is convenient when we wish to define or refer to a large number of sets without writing out all of them explicitly. Using similar notation with the set operations we have defined already, we will be able to "combine" and "operate on" a large

number of sets "all at once". There is really no new mathematical content in this section, but the notation involved in these ideas can be confusing and difficult to work with, at first, so we want to guide you through these ideas carefully.

Relation to Summation Notation

We'll start with a related concept that we have seen before. Remember when we investigated sums of natural numbers in Chapter 1? We mentioned some special notation that allowed us to condense a long string of terms in a sum into one concise form, using the $\sum$ symbol. For instance, we could write an informal sum ("informal" meaning "not rigorous" because of the use of ellipses) in the $\sum$ notation as follows:

$$1 + 2 + 3 + 4 + \cdots + (n-1) + n = \sum_{i=1}^{n} i$$

Why does this notation work and make sense? The **index variable** i is the key component of condensing the sum into this form. Writing "$i = 1$" underneath the $\sum$ symbol means that the value of the variable i should start at 1 and increase by 1 until it reaches the terminal value, n, written above the $\sum$ symbol. For each allowable value of i in that range (from 1 to n), we include a term in the sum of the form written to the right of the $\sum$ symbol; in this case, that term is i. Thus, we should have the terms $1, 2, 3, \ldots, n$ with a + sign between each.

We should point out that it is implicitly understood that writing $i = 1$ and n as the **limits** on the index variable i means i assumes all values that are natural numbers between 1 and n.

Example

Let's see the process of defining indexed sets via an example first. We will also see how to apply set operations to several sets by using an index variable.

Example 3.6.1. We can similarly condense some set operation notation. For instance, let's define the sets $A_1, A_2, A_3, \ldots, A_{10}$ by

$$\begin{aligned} A_1 &= \{1, 2\} \\ A_2 &= \{2, 4\} \\ A_3 &= \{3, 6\} \\ &\vdots \\ A_i &= \{i, 2i\} \\ &\vdots \\ A_{10} &= \{10, 20\} \end{aligned}$$

We included the definition of A_i for an *arbitrary* value i to give these sets a rigorous definition. Without defininining that set—which defines A_i for any relevant value of i—we would be leaving it up to the reader to interpret the pattern among the sets A_1, A_2, A_3, A_{10}, and there could be multiple ways of interpreting that. By defining the term A_i explicitly like this, there is no confusion as to what we want these ten sets to be.

Furthermore, we can more easily express the union of all of these sets, for instance. Remember that the union of two sets is the set containing all elements of both sets (i.e. an element is included in the union if it is in the first set *or* the second set, or possibly both). What is the union of more than two sets? It follows the same idea as the definition for just two sets; we want to include an element in the union if it is in *any* of the constituent sets we are combining via the union operation.

How can we write this union concisely and precisely? Let's follow the same motivation of the $\sum$ notation. The index of these sets runs from 1 to 10, so we should write $i = 1$ below a "$\cup$" symbol and 10 above it. Each term in the union is of the form $\{i, 2i\}$, so we should write that to the right of the "$\cup$" symbol. For *indexed* unions like this one, though, we use a slightly larger "$\bigcup$" symbol, like so:

$$A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{10} = \bigcup_{i=1}^{10} A_i = \bigcup_{i=1}^{10} \{i, 2i\}$$

This is far more concise than writing the elements of all 10 sets, so you can see how *useful* this notation is. We will keep reminding you of the imprecision of the ellipses in the union on the left and tell you that, in fact, an expression like the one on the right is a truly rigorous mathematical statement about this union. The expression on the left is more of an intuitive, heuristic way of describing the union operation applied to these ten sets.

When The Index Set Is Not a Range of Numbers

Let's examine a more difficult example to motivate the next development in this notation technique. What if we asked you to write the following sum in summation notation: the sum of the squared reciprocals of all prime numbers. How can we accomplish this? (Note: We just want to express all the terms of the sum without *evaluating* the sum. That's a difficult endeavor left for another time!)

Unfortunately, we cannot use the exact same notation as above, because we don't want to sum over a range of index values between two natural numbers; rather, we want to only include terms in the sum corresponding to prime numbers. The solution to this is to define an **index set** I that will describe the allowable values of the index that we will then "plug into" the arbitrary term written to the right of the sum.

In this case, if we have a prime number i, we would like to include the term $\frac{1}{i^2}$ in our sum, so this expression will be written to the right of the $\sum$ symbol. We would like to express in our notation that the value i should be a prime

number and that all possible prime numbers should be included. The index set I of allowable values should, therefore, be the set of all prime numbers. That is, we can write this sum as

$$\sum_{i \in I} \frac{1}{i^2} \quad \text{, where } I = \{i \in \mathbb{N} \mid i \text{ is prime}\}$$

Look at what this notation accomplishes! Not only have we condensed an infinite number of terms into one expression, we have indicated that the values of the arbitrary index i should be restricted to prime numbers, which do not behave in the "usual" and convenient way as with a sum like $\sum_{i=1}^{n} i$.

Example 3.6.2. This concept of an *index set* is extremely useful and extends to arbitrary sets and even non-mathematical objects. For instance, in our discussion of sets above we used the set B of all Major League Baseball *teams*. How can we use this set to express the set P of all Major League Baseball *players*? Each team is, itself, a set whose elements are the players on that team, so a union of all of the teams (i.e. a union of all sets in B) should produce exactly this set of all players! In this case, our index set is B, and for each element $b \in B$, we want to include b as a *set* in our union. Thus, we would write

$$P = \bigcup_{b \in B} b$$

The individual terms in this union are not even dependent on natural numbers, so there would have been no way to express this union without the use of an index set, like this. Also, this union is dependent on the fact that the terms of the union are elements of the index set B, but they are also sets themselves; thus, applying the union operation to them makes mathematical sense. This might still seem like an odd idea, so be sure to think carefully about this idea of sets having elements that are sets, themselves.

Reading Indexed Expressions Aloud

To verbalize these types of expressions, and to help you think of them in your head, let's give you an example. We might read the expression up above as

> "The sum, over all i that are prime, of $\frac{1}{i^2}$."
> or
> "The sum of $\frac{1}{i^2}$, where i ranges over all prime numbers."

Likewise, we might read the other expression above as

> "The union, over all b that are MLB teams in the 2012 season, of those b."
> or
> "The union of all sets b, where b ranges over all MLB teams from the 2012 season."

3.6.2 Indexed Unions and Intersections

Let's give a precise definition of this union operation for more than one set, since we have only rigorously defined the union of two sets.

Definition 3.6.3. *The union of a collection of sets A_i indexed by the set I is*

$$\bigcup_{i\in I} A_i = \{x \in U \mid x \in A_i \text{ for some (i.e. at least one) } i \in I\}$$

where we assume there is a set U such that $A_i \subseteq U$ for every $i \in I$.

In mathematical language, the phrase "for some $i \in I$" means that we want there to be *at least one* $i \in I$ with the specified property. If an element x satisfies $x \notin A_i$ for *every* $i \in I$, then this says x is not in any of the sets in our collection, so it should not be included in the union.

Following this idea, we can make a similar definition for set intersections.

Definition 3.6.4. *The intersection of a collection of sets A_i indexed by the set I is*

$$\bigcap_{i\in I} A_i = \{x \in U \mid x \in A_i \text{ for every } i \in I\}$$

where we assume there is a set U such that $A_i \subseteq U$ for every $i \in I$.

3.6.3 Examples

Let's return to a previous example and make these ideas clearer.

Example 3.6.5. Previously, in Example 3.6.1, we defined

$$A_i = \{i, 2i\}$$

for every natural number i between 1 and 10. Another way of defining this collection is to consider the index set $I = [10]$ (recall the notation $[n] = \{i \in \mathbb{N} \mid 1 \leq i \leq n\}$) and define A to be the set

$$A = \{A_i \mid i \in I\} \text{ , where } A_i = \{i, 2i\} \text{ for every } i \in I$$

This defines every set A_i, dependent on the index value i chosen from the index set I, and it "collects" all of these sets into one set A. Then, another way of writing that union we wrote before would be

$$\bigcup_{i\in I} A_i$$

with the given definitions of I and A_i.

(Think carefully about how this union differs from the set A. Also, what exactly is this union? How can we express its elements conveniently? Do we need to list every element? What if we change the index set I to be $\mathbb{N}$? What is the union in that case?)

Example 3.6.6. Let $I = \{1, 2, 3\}$ and, for every $i \in I$, define

$$A_i = \{i - 2, i - 1, i, i + 1, i + 2\}$$

Let's identify and write out the elements of the following sets:

$$\bigcup_{i \in I} A_i \qquad \text{and} \qquad \bigcap_{i \in I} A_i$$

Notice that we can write out the elements of each A_i set, as follows:

$$A_1 = \{-1, 0, 1, 2, 3\}$$
$$A_2 = \{0, 1, 2, 3, 4\}$$
$$A_3 = \{1, 2, 3, 4, 5\}$$

Thus,

$$\bigcup_{i \in I} A_i = A_1 \cup A_2 \cup A_3 = \{-1, 0, 1, 2, 3, 4, 5\}$$

and

$$\bigcap_{i \in I} A_i = A_1 \cap A_2 \cap A_3 = \{1, 2, 3\}$$

Now, consider $J = \{-1, 0, 1\}$, with A_j defined in the same way as before. Let's identify the elements of the sets

$$\bigcup_{j \in J} A_j \qquad \text{and} \qquad \bigcap_{j \in J} A_j$$

Writing out the elements of each set, we can determine that

$$\bigcup_{j \in J} A_j = A_{-1} \cup A_0 \cup A_1 = \{-3, -2, -1, 0, 1, 2, 3\}$$

and

$$\bigcap_{j \in J} A_j = A_{-1} \cap A_0 \cap A_1 = \{-1, 0, 1\}$$

Try answering the same questions with different index sets.

For instance, consider $K = \{1, 2, 3, 4, 5\}$ or $L = \{-3, -2, -1, 0, 1, 2, 3\}$.

Example 3.6.7. Define the index set $I = \mathbb{N}$. For every $i \in I$, define the set

$$C_i = \left\{ x \in \mathbb{R} \mid 1 \leq x \leq \frac{i + 1}{i} \right\}$$

Then we claim that

$$\bigcup_{i \in I} C_i = \{y \in \mathbb{R} \mid 1 \leq y \leq 2\} \quad \text{and} \quad \bigcap_{i \in I} C_i = \{1\}$$

Can you see why these are true? We will discuss the techniques required to prove such equalities later on. For now, we ask you to just think about why these are true. Can you explain them to a classmate or a friend? What sort of techniques might you use to prove these claims?

Example 3.6.8. Let S be the set of students taking this course. For every $s \in S$, let C_s be the set of courses that student s is taking this semester. What do the following expressions represent?

$$\bigcup_{s \in S} C_s \qquad \text{and} \qquad \bigcap_{s \in S} C_s$$

We bet you can identify at least one element of the set on the right!

3.6.4 Partitions

Now that we have a way of writing down a union of many sets, we can define a helpful notion: that of a **partition**. Linguistically speaking, a partition is a way of "breaking something apart into pieces", and that's pretty much what this word means, mathematically speaking.

To wit, a partition is just a collection of subsets of a set that do not overlap and whose union is the entire set. Let's write down that definition here and then see some examples and non-examples. We will have occasion to use this definition many times in the future, so let's get a handle on it now while we're talking about sets and indexed unions.

Definition 3.6.9. *Let A be a set. A* **partition** *of A is a collection of sets that are pairwise disjoint and whose union is A.*

That is, a partition is formed by an index set I and non-empty sets S_i (defined for every $i \in I$) that satisfy the following conditions:

(1) For every $i \in I$, $S_i \subseteq A$.

(2) For every $i, j \in I$ with $i \neq j$, we have $S_i \cap S_j = \varnothing$.

(3) $\displaystyle\bigcup_{i \in I} S_i = A$

The sets S_i are called **parts** *of the partition.*

The idea here is that the sets S_i "carve up" the set A into non-overlapping, nonempty pieces.

Example 3.6.10. Let's see a couple of examples.

(1) Consider the set $\mathbb{N}$. Let O be the set of odd natural numbers, and let E be the set of even natural numbers. Then $\{O, E\}$ is a partition of $\mathbb{N}$. This is because

- $E, O \neq \varnothing$, and
- $E, O \subseteq \mathbb{N}$, and
- $E \cap O = \varnothing$, and
- $E \cup O = \mathbb{N}$

(2) Consider the set $\mathbb{R}$. For every $z \in \mathbb{Z}$, define the set S_z by

$$S_z = \{r \in \mathbb{R} \mid z \leq r < z + 1\}$$

We claim $\{\ldots, S_{-2}, S_{-1}, S_0, S_1, S_2, \ldots\}$ is a partition of $\mathbb{R}$. Can you see why? Try to write out the conditions required for this collection of sets to be a partition, and see if you can understand why they hold.

Specifically, remember that we need these sets to be *pairwise* disjoint. This means that *any* two sets must be disjoint. Notice that this is quite different from requiring the the intersection of *all* the sets together to be empty.

For instance, consider the collection of sets

$$\big\{\, \{1,2\}, \{2,3\}, \{3,4\} \,\big\}$$

This collection is *not* pairwise disjoint because, for example,

$$\{1,2\} \cap \{2,3\} = \{2\} \neq \varnothing$$

However, the intersection of all three sets is empty, because no element is common to all three of them together.

Example 3.6.11. Now, let's see a couple of non-examples.

(1) Consider the set $\mathbb{R}$. Let P be the set of positive real numbers and let N be the set of negative real numbers. Then $\{N, P\}$ is not a partition because $0 \notin N \cup P$.

Can you modify the choices we made here to identify a partition of $\mathbb{R}$ into two parts?

(2) Consider the set $\mathbb{Z}$. Let A_2 be the set of integers that are multiples of 2, let A_3 be the set of integers that are multiples of 3, and let A_5 be the set of integers that are multiples of 5. The collection $\{A_2, A_3, A_5\}$ is not a partition for two reasons.

First, these sets are not pairwise disjoint. Notice that $6 \in A_2$ and $6 \in A_3$, since $6 = 2 \cdot 3$. Second, these sets do not "cover" all of $\mathbb{Z}$. Notice that $7 \in \mathbb{Z}$ but $7 \notin A_2 \cup A_3 \cup A_5$.

As we mentioned, we will have occasion to use this definition frequently in the future, so keep it in mind.

3.6.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is an index set?

(2) Let $I = \mathbb{N}$, and for every $i \in I$, let $A_i = \{i, -i\}$. Why are the following sets all the *same* set?

$$\bigcup_{i \in I} A_i \qquad \bigcup_{x \in \mathbb{N}} A_x \qquad \bigcup_{j \in I} A_j$$

By the way, what are the elements of this set?

(3) List the elements of the following sets:

(a) $\bigcup_{x \in \mathbb{N}} \{x\}$

(b) $\bigcap_{x \in \mathbb{N}} \{x\}$

(c) $\bigcup_{x \in \mathbb{N}} \{x, 0, -x\}$

(4) Why do you think we didn't talk about an "indexed difference" or an "indexed complementation", and only talked about unions and intersections?

(5) What is a partition? What conditions does a collection of sets have to satisfy to be a partition of a set?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Let $A = \{-2, -1, 0, 1, 2\}$. Let $B = \{1, 3, 5\}$.

For every $i \in \mathbb{Z}$, let $S_i = \{i - 2, i, i + 2, i + 4\}$.

What is $\bigcup_{i \in A} S_i$? What is $\bigcap_{x \in B} S_x$?

(2) For every $n \in \mathbb{N}$, let $A_n = [n]$. What is $\bigcap_{n \in \mathbb{N}} A_n$? What about $\bigcup_{n \in \mathbb{N}} A_n$?

(3) Find a way to write the set of all integers between -10 and 10 (inclusive) using set-builder notation. Then, define the same set using an indexed union. Can you do this in a way so that the sets in your union are pairwise-disjoint (meaning that no two of them have any elements in common)? (Hint: Yes, you can.)

(4) For every $n \in \mathbb{N}$, let M_n be the set of all *multiples* of n. (For example, $M_3 = \{3, 6, 9, \dots\}$.) Write a definition for M_n using set-builder notation. Then, express $\mathbb{N}$ as a union, using these sets.

(**Challenge:** Can you use these sets to define a *partition* of $\mathbb{N}$?)

(5) Let X be any set. What is $\bigcup_{S \in \mathcal{P}(X)} S$? What about $\bigcap_{S \in \mathcal{P}(X)} S$?

(It might help to try this with a specific set, like $X = \{1, 2\}$, to see what happens, first.)

(6) Identify an index set and define some sets that allow you to express $\mathbb{Q}$ as an indexed union.

Can you do this so that there are infinitely many elements in the index set?

(**Challenge:** Can you make this collection a *partition* of $\mathbb{Q}$?)

3.7 Cartesian Products

There is one more way of "combining" sets to produce other sets that we want to investigate. This method is based on the idea of **order**. When we define sets by listing the elements, the order is irrelevant; that is, the sets $\{1, 2, 3\}$ and $\{3, 1, 2\}$ and $\{2, 1, 3\}$ are all equal because they contain the same elements. (More specifically, they are all subsets of each other in both directions). Looking at mathematical objects where order *is* relevant, though, allows us to combine sets in new ways and produce new sets.

You are likely already familiar with the idea of the real plane, $\mathbb{R}^2$ (also known as the **Cartesian plane** after the French matheamtician René Descartes). Each "point" on the plane is described by two values, an x-coordinate and a y-coordinate, and the order in which we write those coordinates is important. We usually think of the x-coordinate as first and the y-coordinate as second, and this helps to distinguish two points based on this order. For instance, the point $(1, 0)$ lies on the x-axis but the point $(0, 1)$ lies on the y-axis. They are not the same point.

There is a deeper, mathematical idea underlying the Cartesian plane. Given any two sets, A and B, we can look at the set of all **ordered pairs** of elements from A and B. By **pair** we mean an expression (a, b) where a and b are elements of A and B, respectively. By **ordered** we mean that writing a first and b second is important. In the case of the real plane, it is especially important because any real number could appear as the x-coordinate *or* the y-coordinate of a point, but the point (x, y) is generally different from the point (y, x). (When are they equal? Think carefully about this.)

3.7.1 Definition

Let's give an explicit definition of this new set before examining some examples.

Definition 3.7.1. *Given two sets, A and B, the* **Cartesian product** *of A and B is written as $A \times B$ and defined to be*

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

This definition tells us that the Cartesian product $A \times B$ collects into a new set all of the ordered pairs (a, b), where a is allowed to be any element of A and b is allowed to be any element of B.

Some Technicalities

Notice that we have dropped the assumption of a universal set U. We have discussed some of the issues that arise when we do *not* specify a universal set, but from now on the sets we work with will not address any of these issues. Accordingly, we will only specify a universal set when not doing so would lead to ambiguity.

In the case of this definition, we could specify a universal set by defining the ordered pair (a, b) as a *set*.Specifically, we could define

$$(a, b) = \{\, \{a\}, \{a, b\} \,\}$$

This definition incorporates the *order* of the pair, as well, in the sense that

$$(a, b) = (c, d) \text{ if and only if } a = c \text{ and } b = d$$

Checking the singleton element in the set tells us the first coordinate, and checking the other element in the set with two elements tells us the second coordinate. If we have the ordered pair (a, a), then the set reduces to $\{\{a\}\}$, which tells us a appears in both coordinates.

Using this definition, we could use the universal set $U = \mathcal{P}(\mathcal{P}(A \cup B))$. We won't delve into the technical details of these sets and definitions, but we thought it prudent to point out that such definitions exist. The important point to remember from this section is given above:

> Two ordered pairs are equal if and only if *both* of their coordinates are equal.

This is why we call them **ordered pairs**.

3.7.2 Examples

The Cartesian plane is $\mathbb{R} \times \mathbb{R}$, which is why we sometimes see this written as $\mathbb{R}^2$. Indeed, if $A = B$, then we sometimes write the Cartesian product as $A \times A = A^2$, if there is no confusion about the fact that A is a set (and not a number). Let's see some more examples where the two sets in the Cartesian product are not the same.

Example 3.7.2. Define the sets $A = \{a, b, c\}$ and $B = \{6, 7\}$ and $C = \{b, c, d\}$. Then we can list the elements of the following Cartesian products:

$$A \times B = \{(a, 6), (a, 7), (b, 6), (b, 7), (c, 6), (c, 7)\}$$
$$B \times C = \{(6, b), (6, c), (6, d), (7, b), (7, c), (7, d)\}$$
$$A \times C = \{(a, b), (a, c), (a, d), (b, b), (b, c), (b, d), (c, b), (c, c), (c, d)\}$$
$$C \times B = \{(b, 6), (b, 7), (c, 6), (c, 7), (d, 6), (d, 7)\}$$

Notice that, in general, $B \times C \neq C \times B$, as this example shows.

(Can you identify the situations where $A \times B = B \times A$? What conditions must we impose on the sets A and B to make this equality true?)

Ordered Triples and Beyond

This idea also extends to Cartesian products of three or more sets. We simply write ordered *triples* for a Cartesian product of three sets and, in general, for the Cartesian product of n sets, we write ordered n-tuples. (Again, we point out that there are set-theoretic ways of defining these ordered n-tuples, but we will not investigate those details.)

Example 3.7.3. The Cartesian product $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ (sometimes written as $\mathbb{N}^3$) is the set of all ordered triples of natural numbers. For instance, $(1, 2, 3) \in \mathbb{N}^3$ and $(7, 7, 100) \in \mathbb{N}^3$, but $(0, 1, 2) \notin \mathbb{N}^3$ and $(1, 2, 3, 4) \notin \mathbb{N}^3$.

Notice the subtle distinction between $\mathbb{N}^3$ and $(\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$. A typical element of $\mathbb{N}^3$ is an ordered *triple* whose coordinates are each a natural number. A typical element of $(\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$ is an ordered *pair*, the first coordinate of which is also an ordered pair (of natural numbers) and the second coordinate of which is a natural number. That is, $((1, 2), 3) \in (\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$ but $((1, 2), 3) \notin \mathbb{N}^3$. This shows the two are *different sets*.

There is, however, a *natural* way of relating the two sets which essentially "drops the parentheses" around the first coordinate (the ordered pair). We will discuss this later on when we examine functions and *bijections*. For now, though, we just want you to notice the subtle differences between the two sets and remember that a Cartesian product of two sets is a *set* of *ordered pairs*, where each coordinate is drawn from the corresponding constituent set.

Example 3.7.4. What happens if $B = \varnothing$, say? Look back at the definition of $A \times B$. There are actually *no* elements of B to write as the second "coordinate" of the ordered pair, so we actually have no elements of $A \times B$ to include! Thus,

$$A \times \varnothing = \varnothing$$

for any set A. Similarly, $\varnothing \times B = \varnothing$, for any set B.

3.7.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between $\mathbb{R} \times \mathbb{N}$ and $\mathbb{N} \times \mathbb{R}$? Give an example of an ordered pair that is an element of one set, but not the other. Then, give an example of an ordered pair that is actually an element of *both* sets.

(2) What is $\varnothing \times \mathbb{Z}$?

(3) Write out all the elements of the set $\{\heartsuit, \diamondsuit\} \times \{\circledcirc, \square, \heartsuit\}$.

(4) What is the difference between $(\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$ and $\mathbb{N} \times (\mathbb{N} \times \mathbb{N})$? Why aren't they technically the same set? Can you explain why they are "essentially" the same set?

(5) Let A, B, C be sets. Suppose $A \subseteq B$. Do you think $A \times C \subseteq B \times C$ is true? Why or why not?

(6) Give an example of a set S such that $\left(\frac{1}{2}, -1\right) \in S$.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Write out the elements of $[3] \times [2]$.

Can you make a conjecture about how many elements $[m] \times [n]$ has, for any $m, n \in \mathbb{N}$? (How might you try to prove your conjecture?)

(2) Give an example of an element of the set $\mathbb{N} \times \mathcal{P}(\mathbb{Z})$.

(3) Give an example of an element of the set $\big((\mathbb{R} \times \mathbb{N}) \times \mathbb{Q}\big) \cap \big((\mathbb{Q} \times \mathbb{Z}) \times \mathbb{N}\big)$.

(4) Give an example of sets C, D such that $C \times D = D \times C$.

Follow-up challenge: Can you characterize *all possible situations* like this one. What must be true about C and D? Can you prove it?

(5) Write out the elements of $\mathcal{P}\big([1] \times [2]\big)$.

(6) For every $n \in \mathbb{N}$, let $A_n = [n] \times [n]$. Consider the set

$$B = \bigcup_{n \in \mathbb{N}} A_n$$

Is $B = \mathbb{N} \times \mathbb{N}$ or not? Explain, with examples.

(7) If you know some simple computer programming, try writing code (in your favorite language) that will input $m, n \in \mathbb{N}$ and print out all the elements of $[m] \times [n]$. (Use some pseudocode if you're not totally comfortable with programming.) How long do you think this takes to run, depending on m and n?

3.8 [Optional Reading] Defining the Set of Natural Numbers

Our goal in this section is to put the natural numbers $\mathbb{N}$ on a rigorous, mathematical foundation. Specifically, we will prove that the natural numbers exist, by defining and deriving them from the axioms and principles of set theory. We will then discuss a few of their properties. We will use some of those properties to define and prove the Principle of Mathematical Induction in Chapter 5, after discussing some basic principles and results of mathematical logic.

3.8.1 Definition

How do we *define* the natural numbers in terms of sets? We intuitively know what they are. We start with 1 and repeatedly add 1, obtaining all of the other natural numbers. Thus, we have to identify what we mean by "1" and what we mean by "add 1", in terms of sets. To do this, let's start by thinking about 0. We stated before that we will not include 0 in the set $\mathbb{N}$, but some authors do, and it will aid us in deriving $\mathbb{N}$, right now, to consider it. We know of exactly one set that contains no elements, the empty set. Thus, it makes sense to *associate* 0 with the empty set; in fact, we *define* $0 = \varnothing$. Next, we wish to define 1, and following our definition of 0, it makes sense to choose a set that has exactly one element. (A set with one element is also known as a **singleton**.) There are several such sets:

$$\{\varnothing\}, \{\{\varnothing\}\}, \{\{\varnothing, \{\varnothing\}\}\}$$

How do we choose a *representative* singleton to represent 1? Keeping in mind that we want to continue this process and eventually define 2 (and 3, and so on) in terms of previous numbers, it makes sense now to define 1 in terms of the only object we have at our disposal: 0. Thus, let us *choose* to define

$$1 = \{0\} = \{\varnothing\}$$

This guarantees $0 \neq 1$.

Next to define 2, we consider sets containing two elements, like

$$\{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\{\varnothing\}\}\}, \{\{\varnothing\}, \{\{\varnothing\}\}\}$$

and so on. We seek a natural representative, and we notice that the first set listed above contains the two objects, 0 and 1, that we have already defined! Thus, defining $2 = \{0, 1\}$ is a natural choice and, again, we know $2 \neq 0$ and $2 \neq 1$.

Successors

This gives us an intuitive idea of how to continue this process and define any natural number: for any $n \in \mathbb{N}$, we define

$$n = \{0, 1, 2, \ldots, n-2, n-1\}$$

However, given a set, it would be quite difficult to verify, using this definition, whether or not that set represents a natural number. We would like a *better* definition of the elements of $\mathbb{N}$; we want to know, given any set, whether it belongs in $\mathbb{N}$. Look back at the element n above; we could also write

$$n = \{0, 1, 2, \ldots, n-2, n-1\} = \{0, 1, 2, \ldots, n-2\} \cup \{n-1\} = (n-1) \cup \{n-1\}$$

Look at that! We have a natural way of defining an element of $\mathbb{N}$ in terms of the previous element and in terms of set operations. This motivates the following definition.

Definition 3.8.1. *Given any set X, the* **successor** *of X, denoted by $S(X)$, is defined to be $S(X) = X \cup \{X\}$.*

This definition applies to all sets, but in the context of natural numbers, it means that the successor of n is precisely that natural number we "know" intuitively to be one larger, namely $n + 1$.

Inductive Sets

This brings us closer to our definition of $\mathbb{N}$. We certainly want $1 \in \mathbb{N}$ and we also want $S(n) \in \mathbb{N}$ for any element $n \in \mathbb{N}$. To codify this symbolically, we make the following definitions:

Definition 3.8.2. *A set I is called* **inductive** *provided*

1. $1 \in I$
2. *If $n \in I$, then $S(n) \in I$, as well.*

Certainly, $\mathbb{N}$ itself (as we hope to define it) should be an inductive set. Are there *other* inductive sets? Think about this. What properties would they have? Would they contain elements that are *not* natural numbers? We don't want to address these questions in depth, but for the sake of our discussion here, we will point out that there are, indeed, other inductive sets. We don't want any of those sets to be $\mathbb{N}$, so we make this definition:

Definition 3.8.3. *The set of all* ***natural numbers*** *is the set*

$$\mathbb{N} := \{x \mid \text{ for every inductive set } I,\ x \in I\}$$

Put another way, $\mathbb{N}$ *is the* smallest *inductive set, in the sense of set inclusion:*

$$\mathbb{N} = \bigcap_{I \in \{S \mid S \text{ is inductive}\}} I$$

This dictates that $\mathbb{N}$ *is a subset of every inductive set.*

This gives us the "checking property" we desired. Any set x is a natural number (i.e. $x \in \mathbb{N}$) if and only if it is an element of *every* inductive set (i.e. $x \in I$ for every inductive set I). This also tells us that $\mathbb{N} \subseteq I$ for every inductive set I.

There are some other set theoretic discussions that could be made here: How do we know that such an infinite set exists? (In actuality, we need to make this an *axiom* of set theory! Assuming these types of sets exist, how do we characterize those other inductive sets that are not $\mathbb{N}$? Addressing these questions lies outside the scope and goals of this course, so we will not address them. We will, however, mention a few properties of $\mathbb{N}$ now, specifically ones that will be useful in setting mathematical induction on a rigorous foundation. (In case you're wondering, think about the set of integers, $\mathbb{Z}$. Try to explain why this set is, indeed, inductive. What about $\mathbb{R}$? What about $\mathbb{Z} - \mathbb{N}$?)

Properties of $\mathbb{N}$

Before we define the principle of induction, let's think about some of the common properties and uses of natural numbers: orderings and arithmetic. Given any two natural numbers, we can *compare* them and decide which one is larger and which one is smaller (or if they are equal). We usually write this with symbols like $1 < 3$, $1 \leq 5$, $4 \not< 2$, $3 = 3$, etc.

Can we phrase these comparisons in terms of *sets*, knowing that we have now defined the elements of $\mathbb{N}$ as sets, themselves? Yes, we can! Look back at the definition of *successor*. Built into that definition is the fact that $X \in S(X)$! This observation gives us the following definition:

Definition 3.8.4. *Given two natural numbers* $m, n \in \mathbb{N}$*, we write* $m < n$ *if and only if* $m \in n$*.*

This defines an *order relation* on the set $\mathbb{N}$. We will discuss the concepts of relations and orders later on in the book (in Section 6.3).

What about arithmetic? What is $m + n$ in terms of the sets m and n? How do we define this operation and its output? How do we know $m + n$ is another natural number? Can we be sure that $m + n = n + m$? These are questions we can address later on after discussing functions and relations.

3.8.2 Principle of Mathematical Induction

For now, let us present a more rigorous version of induction:

Theorem 3.8.5 (Principle of Mathematical Induction)**.** *Let $P(n)$ be some "fact" or "observation" that depends on the natural number n. Assume that*

1. *$P(1)$ is a true statement.*
2. *Given any $k \in \mathbb{N}$, if $P(k)$ is true, then we can conclude* necessarily *that $P(k+1)$ is true.*

Then the statement $P(n)$ must be true for every *natural number $n \in \mathbb{N}$.*

Let us first prove this theorem before discussing its assumptions and consequences in detail.

Proof. Define the set S to be the natural numbers for which the statement P is true. That is, define $S = \{n \in \mathbb{N} \mid P(n) \text{ is true}\}$. By definition, $S \subseteq \mathbb{N}$.

Furthermore, the assumptions of the theorem guarantee that $1 \in S$ and that whenever $k \in S$, we know $k+1 \in S$, as well. This means S is an *inductive* set. By the observation we made after defining $\mathbb{N}$, we know that $\mathbb{N} \subseteq S$.

Therefore $S = \mathbb{N}$, so the statement $P(n)$ is true for every natural number n. □

This is pretty *slick*, right? It seems that all of the desired conclusions "fell out of" our definitions! In this sense, the definitions and axioms are *natural* choices, because they accomplish what our *intuition* already "knew" about the set $\mathbb{N}$ and its properties.

There are a few minor issues that we have left undiscussed. Specifically, what do we mean by a "fact" or "observation" that *depends* on a natural number n? What does it mean to *necessarily conclude* that $P(k+1)$ is true when $P(k)$ is true? What do we even mean by *true*? These are all deep mathematical questions and involve a thorough study of logic, and we will discuss these issues in the next chapter! Onward! Huzzah!

3.8.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is an inductive set? Give an example of one that is not $\mathbb{N}$ or $\mathbb{Z}$.

(2) We defined $S = \{n \in \mathbb{N} \mid P(n) \text{ is true}\}$ in the proof of the Principle of Mathematical Induction. What does this mean? Describe this set in words.

(3) Come up with your own analogy for how Induction works.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) What if we changed the definition of **successor** to be $S(X) = \{X\}$. Using $0 = \varnothing$, what are 1, 2, 3, and 4 in terms of sets? Do they still satisfy the equality $n = \{0, 1, \ldots, n-1\}$? If not, do they satisfy some other relationship? Explore!

(2) Have a debate with a friend about whether or not infinite sets exist. Why do we need to *assume* the existence of an inductive set to define $\mathbb{N}$? Does this seem valid to you? Does it make sense, physically? Mathematically?

(3) Consider a simple arithmetic statement, like $1 + 2 = 3$. Write out the numbers 1, 2, and 3 in terms of sets, and see how this equation might make sense. What does "+" mean, in this context?

(4) Investigate how one might define $\mathbb{Z}$, using $\mathbb{N}$. Do some exploring online or in books, or make up an idea on your own.

3.9 Proofs Involving Sets

Now that we've gone through many definitions and examples, to introduce what sets are and how to manipulate them, let's actually write up some rigorous, mathematically correct, and well-written **proofs** about sets. All of the propositions/lemmas contained here are useful facts that we can cite later on, and we would expect you to be able to prove claims like these. (Note: A lemma is just a small result that requires some proof, and can be cited later to prove more significant theorems.) Furthermore, all of these proofs are of the type of quality and rigor that we will expect from you in the future, the very near future . . . Use these as guidelines, if you'd like!

3.9.1 Logic and Rigor: Using Definitions

The main point we'd like to emphasize here—as we transition from descriptive, "wordy", and intuitive proofs into more rigorous, mathematically correct, and formally-written ones—is that **formal definitions are very important**. Fundamentally, they're essential because when we say, for example, "$A \cup B$", we need to know that you know exactly what that symbol means and how it operates on the sets A and B.

As another example, when we say "Prove $A = B$", we have a very specific goal in mind, and you need to be on the same page. It always helps to have an intuitive understanding of the main concepts—"Oh, the statement $A = B$ just

means that A and B have the same elements in them"—but this is *not* the type of language/ideas we want to use in a rigorous proof. To prove a statement like $A = B$, we need you to **appeal to the definition** of "=" in the context of sets: $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

This is what we mean when we say "satisfy a definition" or "appeal to a definition": to prove that some mathematical object has a certain property, you must demonstrate that the object satisfies the formal definition of that property. If you aren't familiar with that definition, or have forgotten how to state it precisely ... by all means, go look it up! We realize this is a lot of new information to ingest, and there's nothing wrong with forgetting something when it's still new to you. By doing this, you'll start to internalize these ideas more quickly and more solidly.

You'll see how we use the definitions of, for instance, "$\subseteq$" and "=" and "$\cap$" and so on in the examples below. For each proposition/lemma, we will end up writing a formal proof, but we will also write a little bit about how we would approach *coming up with* such a proof. Oftentimes this is the difficult part! We think you'll notice that many of those explanations will amount to just recalling a relevant definition and thinking about what it means and how it applies in the given situation. In a way, that's what a lot of mathematics is. We just allow our definitions we use to get more and more complicated.

3.9.2 Proving "$\subseteq$"

Recall the definition of **subset**, because we will use it frequently here:

Definition 3.9.1. *Given two sets A and B, if every element of A is* also *an element of B, then we say A is a* **subset** *of B.*

Say we are presented with the following problem:

> Let A be the set ... and let B be the set ... Prove that $A \subseteq B$.

How can we satisfy the definition of $A \subseteq B$ to prove this claim? Yes, the intuitive idea is that "every element of A is also an element of B", but we shouldn't just try to dance around the issue and try to make that sentence our conclusion. Rather, we need to verify that *every* element of A is also *necesarily* an element of B. This is where the wonderful phrase **"arbitrary and fixed"** will come in handy!

The Phrase "Arbitrary and Fixed"

How can we talk about *all possible elements* of A all at once? Of course, we might not need to do this if A has only, say, 3 elements; then, we can just work with them one by one. But what if A has 100 elements? Or 1 million? Or *infinitely* many? How can we prove something about all of them at once in a reasonable way?

What we will do is introduce an **arbitrary and fixed** element of A so we have something to work with. This element will be **arbitrary** in the sense that

we make no extra assumptions about what it is or what properties it has, only that it is an element of A. This element will be **fixed** in the sense that we will assign it some variable name (usually a letter, like a or x or t or something) and this letter will represent the *same* object throughout the remainder of our proof. As long as we can prove our goal for this element, then we will have simultaneously proven something about *all* elements of A. Voilà!

Examples

Let's see this process in action to really get the point across. We'll begin with the statement to be proven, then describe our thought processes in coming up with a proof, and then present our formal, written proof.

Lemma 3.9.2. *Let A, B, X be any sets.*

If $X \subseteq A$ and $X \subseteq B$, then $X \subseteq A \cap B$.

Intuition: Consider drawing a Venn diagram to represent this situation. To make sure the assumptions $X \subseteq A$ and $X \subseteq B$ both hold true, we need to make the set X "lie inside" both A and B. Accordingly, this means X must lie entirely "inside" where A and B overlap, which is what $A \cap B$ represents. This helps us realize that this statement is, indeed, True. But it doesn't prove anything!

To *prove* the statement, we will introduce an arbitrary and fixed element $x \in X$. What do we know about it? Well, we assumed that $X \subseteq A$. The definition of "$\subseteq$" means that any element of X is *also* an element of A. But we know x is an element of X; this means it is also an element of A. How convenient! We can make some similar statements about x and X and B that will tell us $x \in B$. What does this mean, overall? Oh hey, the definition of "$\cap$" applies, and tells us $x \in A \cap B$. Brilliant! Now, let's write it up.

Proof. Let $x \in X$ be arbitrary and fixed.

By assumption $X \subseteq A$, so $x \in A$, as well, by the definition of $\subseteq$.

Similarly, by assumption $X \subseteq B$, so $x \in B$, as well.

Since $x \in A$ *and* $x \in B$, this means that $x \in A \cap B$, by the definition of $\cap$.

Overall, we have shown that whenever $x \in X$, it is also true that $x \in A \cap B$. Since $x \in X$ was arbitrary, we conclude that $X \subseteq A \cap B$. □

Not so bad, right? Let's try another one, a slightly harder one, even.

Proposition 3.9.3. *Let A and B be any sets. Then, $\mathcal{P}(A) \cap \mathcal{P}(B) \subseteq \mathcal{P}(A \cap B)$.*

Whoa, is this really true? Look back at Problem 6 in Section 3.5, and you'll see an example. This claim states that it is true, *in general*, and not just for that example. Let's figure out why, and then prove it.

Intuition: There are several layers of definitions at work here. In particular, the *power set* operation might be confusing to you. The key is to just remember

that definition: $\mathcal{P}(A)$ is the set of all *subsets of* A. Now, the main claim here is one of a *subset* relationship: whatever the set $\mathcal{P}(A) \cap \mathcal{P}(B)$ is (we'll analyze it later, but it's important that you recognize immediately what *type* of object it is: a *set*), it is supposed to be a subset of whatever the set $\mathcal{P}(A \cap B)$ is. That's it, and it's important to notice that this is really motivates the overarching form of the forthcoming proof.

Without even having to think about what $\mathcal{P}(A) \cap \mathcal{P}(B)$ means, we can be sure that our proof will start with "Let $X \in \mathcal{P}(A) \cap \mathcal{P}(B)$ be arbitrary and fixed". This is because we need to satisfy the definition of "$\subseteq$" by taking any old element of the set on the left and deducing that it is also an element of the set on the right. This is what we mean by the **structure** of the proof.

What does an element $X \in \mathcal{P}(A) \cap \mathcal{P}(B)$ "look like"? It's a set, and it's an element of both $\mathcal{P}(A)$ and $\mathcal{P}(B)$. This means ... well, we're actually going to skip ahead and jump right into the formal proof now, because we'll just find ourselves repeating the same words down below anyway. But before going ahead to read *ours*, we think you should go off and try to write your *own* proof. Then, when you're done, you can compare and see whether you are correct, whether it has the same steps as ours, whether it's written clearly, and so on. See what you can do!

Proof. Let $X \in \mathcal{P}(A) \cap \mathcal{P}(B)$ be arbitrary and fixed.

By the definition of $\cap$, this means $X \in \mathcal{P}(A)$ *and* $X \in \mathcal{P}(B)$.

Since $X \in \mathcal{P}(A)$, we know $X \subseteq A$, by the definition of power set.

Similarly, since $X \in \mathcal{P}(B)$, we know $X \subseteq B$.

Since $X \subseteq A$ and $X \subseteq B$, we know that $X \subseteq A \cap B$ by Lemma 3.9.2 that we just proved.

Now, since $X \subseteq A \cap B$, we know $X \in \mathcal{P}(A \cap B)$, by the definition of power set.

Since X was arbitrary, we conclude that $\mathcal{P}(A) \cap \mathcal{P}(B) \subseteq \mathcal{P}(A \cap B)$. □

Did you do what we did? Did you also cite the previous lemma? Did you instead prove that result all over again without realizing it? Consider that a lesson learned! One of the major benefits of proving results is that we get to use them in later proofs! There's nothing technically wrong with proving the previous result again in the middle of this proof; it just might save a little bit of time and writing to not do so. If you find yourself working on a problem and thinking, "Hey, this feels familiar ...", go back and look for related theorems or lemmas or examples. Maybe you can use some already-acquired knowledge to your advantage.

3.9.3 Proving "="

Double-Containment Proofs

Again, we will need to recall the definition of "=" (in the context of sets), since we will be using it frequently here.

Definition 3.9.4. *We say two sets, A and B, are* **equal***, and write $A = B$, if and only if $A \subseteq B$ and $B \subseteq A$.*

That's it! It's completely built up from a previous definition, that of "$\subseteq$" (since that of "$\supseteq$" is completely equivalent). Thus, this isn't really a new technique, per se, because it's really a repeated application of a previous technique. That is to say, to prove $A = B$, we just need to use the technique used in the last subsection and prove $A \subseteq B$ and then prove $B \subseteq A$.

This technique is so common, in fact, that it is given a name: **double-containment**. When we prove two sets are subsets of each other, both ways, and conclude that they are equal, we call this a **double-containment proof**.

Examples

Let's see an example of this double-containment technique in action.

Lemma 3.9.5. *Let A and B be any sets. Then, $A - (A \cap B) = A - B$.*

Intuition: As usual, we could draw a Venn diagram to convince ourselves of this truth, but that doesn't prove anything. Instead, we will follow a double-containment proof. If we take an element $x \in A - (A \cap B)$, we can apply the definition of "$-$" first, and then "$\cap$", to deduce something about x. Hopefully, it will tell us that $x \in A - B$. Then, if we take an element $y \in A - B$, we can apply some definitions to hopefully deduce that $x \in A - (A \cap B)$. Maybe we aren't sure yet exactly how to do so, but by looking at that Venn diagram and using the definitions, we can surely figure it out. Why don't you try to do it first, then read our proof!

Proof. We will show $A - (A \cap B) = A - B$ by a double-containment proof.

("$\subseteq$") First, let $x \in A - (A \cap B)$ be arbitrary and fixed. By the definition of "$-$", we know that $x \in A$ and $x \notin A \cap B$. This means it is *not* true that x is an element of *both* A and B. We already know $x \in A$, so we deduce that $x \notin B$.

Thus, $x \in A$ and $x \notin B$, so by the definition of "$-$", we deduce that $x \in A - B$. This shows $A - (A \cap B) \subseteq A - B$.

("$\supseteq$") Second, let $y \in A - B$ be arbitrary and fixed. By the definition of "$-$", this means $y \in A$ and $y \notin B$. Now, since y is not an element of B, this means that certainly y is not an element of *both* A and B. That is, $y \notin A \cap B$, by the definition of "$\cap$".

Since we know $y \in A$ and $y \notin A \cap B$, we deduce that $y \in A - (A \cap B)$. This

shows $A - B \subseteq A - (A \cap B)$.

Overall, a double-containment proof has shown that $A - (A \cap B) = A - B$. $\square$

Look at the overall structure of this proof. We knew there would be two parts to it, since it is a double-containment proof, but we were also kind enough to *point this out ahead of time* for our intrepid reader, and separate those two sections appropriately. It would be technically correct to ignore this and just dive right in to the proof, but this might leave a reader confused. The whole point of a proof is to *convince someone else* of a truth that you have already figured out, so you might as well make it as easy as possible for them to follow what you're doing.

Let's see another example of proving two sets are equal. This one will be a little different, because one of the parts of the double-containment will make use of the complement operation. As a preview, spend a minute now thinking about why the statements $A \subseteq B$ and $\overline{B} \subseteq \overline{A}$ are *equivalent* (supposing there is some universal set $A, B \subseteq U$). Draw a Venn diagram and try some examples. Try to prove it, even!

Proposition 3.9.6.

$$\left\{ x \in \mathbb{N} \mid x + \frac{8}{x} \leq 6 \right\} = \{2, 3, 4\}$$

Proof. Let's define $A = \left\{ x \in \mathbb{N} \mid x + \frac{8}{x} \leq 6 \right\}$, and $B = \{2, 3, 4\}$.

To show $A = B$, we will show $A \subseteq B$ and $B \subseteq A$.

First, we will show $B \subseteq A$. We can consider each of the three elements separately, and verify that they satisfy the defining inequality of B:

$$2 + \frac{8}{2} = 6 \leq 6$$
$$3 + \frac{8}{3} = \frac{17}{3} \leq 6$$
$$4 + \frac{8}{4} = 6 \leq 6$$

Since $2, 3, 4 \in \mathbb{N}$, we deduce that $2 \in A$ and $3 \in A$ and $4 \in A$, so $B \subseteq A$.

Next, to show $A \subseteq B$, we will show that $\overline{B} \subseteq \overline{A}$, where the complement is taken in the context of $\mathbb{N}$. That is, we will show that all of the natural numbers $1, 5, 6, 7, \ldots$ are *not* elements of A.

To show this, we will verify that the defining inequality of A is *not* satisfied by any of those elements.

The first two cases can be considered easily: $1 + \frac{8}{1} = 9 \not\leq 6$ and $5 + \frac{8}{5} = \frac{33}{5} \not\leq 6$.

To consider the other cases, we can take an arbitrary and fixed $x \in \mathbb{N}$ with $x \geq 6$. Notice, then, that $x + \frac{8}{x} \geq 6 + \frac{8}{x} > 6$, since $\frac{8}{x} > 0$.

This shows that *only* 2,3,4 satisfy the defining inequality of A.

Overall, by a double-containment argument, we have proven that $A = B$. $\square$

Think carefully, again, about why the method employed in the second half of the proof is valid. (It is actually an instance of using the **contrapositive** of a conditional statement, but we haven't yet defined any of those terms; we will do so in the next chapter on logic.)

Let's see another example of proving set equality. This one is only slightly different in that we are proving some set is actually the *empty set* and, to do so, we will prove that it has *no elements*.

Proposition 3.9.7. *For every $n \in \mathbb{N}$, define $S_n = \mathbb{N} - [n]$. Then*

$$\bigcap_{n \in \mathbb{N}} S_n = \varnothing$$

We suggest you play around with this statement first, if it doesn't make sense. For instance, try identifying the element of the sets S_1, and $S_1 \cap S_2$, and $S_1 \cap S_2 \cap S_3$, and so on. Try to come up with a candidate element of the big intersection on the left, and then figure out why it actually is *not* an element of that set. After that, try to figure out a formal proof and write it out; then, look at ours below!

Proof. Let's define $T = \bigcap_{n \in \mathbb{N}} S_n$, so we can refer to it later.

To prove $T = \varnothing$, we will show that T does not contain any elements. Notice that T is formed by an intersection of many sets of natural numbers, so it's clear that the only *possibilities* for elements of T are natural numbers.

Consider an arbitrary and fixed $x \in \mathbb{N}$. We want to show that $x \notin T$.

Observe that $x \in [x] = \{1, 2, \ldots, x\}$. Thus, $x \notin \mathbb{N} - [x]$, by the definition of "$-$".

By definition, T contains the elements that belong to all of the sets of the form $\mathbb{N} - [n]$. We have identified (at least) one set, $\mathbb{N} - [x]$, of the intersection such that x does *not* belong to that set. Accordingly, x cannot be an element of T, since it does not belong to *all* such sets. Therefore, $x \notin T$.

Since $x \in \mathbb{N}$ was arbitrary, we have proven that T contains no natural numbers as elements, and therefore it has *no* elements at all. $\square$

Summary: Let's make one more statement about why this technique works. We showed that there are no elements of T, i.e. that $T \subseteq \varnothing$. This completes the entire process, because it is trivially true, as well, that $\varnothing \subseteq T$. (This claim holds for any set.) Thus, one of the parts of the double-containment argument is already achieved, and we can conclude $T = \varnothing$.

Alright, one more example. We want to include this one because it gives us further practice in working with indexed set operations. You will find many

similar problems in the exercises for this section. We encourage you to work on as many of them as you can!

Proposition 3.9.8. *For every $n \in \mathbb{N}$, define $A_n = \left\{x \in \mathbb{R} \mid 0 \leq x < \frac{1}{n}\right\}$. Then,*

$$\bigcap_{n\in\mathbb{N}} A_n = \{0\}$$

Think about what this claim means. Draw a picture of the A_n sets on a number line. What does the "$\bigcap$" intersection accomplish? Why does it work out that 0 is an element of that intersection? Why is it the *only* element?

The definition of $\bigcap$ will be crucial in this proof, so let's recall the definition here. The key phrase is *for every*:

Definition 3.9.9. *The intersection of a collection of sets A_i indexed by the set I is*

$$\bigcap_{i\in I} A_i = \{x \in U \mid x \in A_i \text{ for every } i \in I\}$$

where we assume there is a set U such that $A_i \subseteq U$ for every $i \in I$.

That is, remember that the indexed intersection of several sets collects together the elements that belong to *all* of the constituent sets. Thus, in our proof below, you will see that we need to prove that (1) 0 is, indeed, an element of *all* of the A_n sets, and (2) *no other* number is an element of all of them, i.e. for every nonzero real number, we can identify at least one of the A_n sets such that the number is not an element of that set.

Proof. First, we will prove that

$$\{0\} \subseteq \bigcap_{n\in\mathbb{N}} A_n$$

This requires us to show that $0 \in A_n$ for *every* $n \in \mathbb{N}$.

Let $n \in \mathbb{N}$ be arbitrary and fixed. Notice that the inequality $0 \leq 0 < \frac{1}{n}$ does, indeed hold.

(Note: You might be worried because "in the limit" 0 is not less than every fraction $\frac{1}{n}$ "all at once", but that is not the point! Think of it this way: Is $0 \in A_1$? Yes, $0 \leq 0 < 1$. Is $0 \in A_2$? Yes, $0 \leq 0 < \frac{1}{2}$. Is $0 \in A_3$? Yes, $0 \leq 0 < \frac{1}{3}$. And so on. The inequality holds for every $n \in \mathbb{N}$ *individually*, so 0 is an element of *every* such set. If you weren't worried about this, never mind! Move right along!)

Thus, $0 \in A_n$ for every $n \in \mathbb{N}$, and so $0 \in \bigcap_{n\in\mathbb{N}} A_n$, by the definition of "$\bigcap$". This shows that $\{0\} \subseteq \bigcap_{n\in\mathbb{N}} A_n$.

Second, we will prove that

$$\bigcap_{n\in\mathbb{N}} A_n \subseteq \{0\}$$

We will do this by considering the *complements* of these sets, in the context of $\mathbb{R}$. Specifically, we will show that

$$\overline{\{0\}} \subseteq \overline{\bigcap_{n\in\mathbb{N}} A_n}$$

which means we will show that every nonzero real number is *not* an element of *every* A_n.

Let $x \in \mathbb{R}$ be arbitrary and fixed, with the property that $x \neq 0$. Either $x > 0$ or $x < 0$, then, so let's consider each case separately.

Case 1: Suppose $x > 0$. Consider the number $\frac{1}{x} \in \mathbb{R}$. Since the natural numbers are infinite and unbounded in $\mathbb{R}$, we can choose a natural number M that is *bigger* than that real number. That is, we can choose $M \in \mathbb{N}$ such that $M > \frac{1}{x}$.

(Note: Think about why this works. We haven't *proven* that $\mathbb{N}$ is infinite, or that the numbers "go on forever" along the number line of $\mathbb{R}$, but we hope these ideas make sense to you, intuitively.)

Take such an $M \in \mathbb{N}$ with $M > \frac{1}{x}$. Since $x > 0$, we can multiply the inequality on both sides by x; since $M > 0$ (so $\frac{1}{M} > 0$) we can then multiply again by $\frac{1}{M}$. This yields $x > \frac{1}{M}$. Accordingly $x \notin A_M$, because $-\frac{1}{M} < x < \frac{1}{M}$ is False.

Since $x \notin A_M$, then surely x is not an element of *all* such sets. Therefore, $x \notin \bigcap_{n\in\mathbb{N}} A_n$.

Case 2: Next, suppose that $x < 0$. We will make a similar argument as the previous case; this time, we will just consider $-x$, since $-x > 0$. Using the same logic as above, we can surely identify a natural number $M \in \mathbb{N}$ that satisfies $M > \frac{1}{-x} = -\frac{1}{x}$. Manipulating the inequality tells us that $x < -\frac{1}{M}$. Thus, $x \notin A_M$, and so $x \notin \bigcap_{n\in\mathbb{N}} A_n$.

Therefore, we have shown that any $x \in \mathbb{R}$ with $x \neq 0$ is not an element of at least one of the A_n sets, so any such x is not an element of their intersection. Thus, $\{0\} \subseteq \bigcap_{n\in\mathbb{N}} A_n$, and we have proven the claim by a double-containment argument. □

This proof is harder than the other ones, we think, so make sure to read it a couple times to make sure you see what happens in every step. In particular, think about how we came up with the step where we chose $M \in \mathbb{N}$ that satisfies $M > \frac{1}{x}$. Do you think we magically intuited that choice? Or do we think we recognized that we wanted $x < \frac{1}{M}$ to be true for some M, and manipulated the inequality backwards to figure out how to make that happen?

3.9.4 Disproving Claims

Motivating Example

Consider the following proposed claim:

> For any sets F, G, H, if $F \subseteq G \cup H$, then either $F \subseteq G$ or $F \subseteq H$.

Is this claim True? If so, how would we prove it? Well, we'd take an arbitrary and fixed element $x \in F$. Since $F \subseteq G \cup H$, this would tell us $x \in G \cup H$, as well. Accordingly, either $x \in G$ or $x \in H$. Is that it? Are we done with the proof?

We hope you recognize that this does *not* work! In particular, we have not satisfied the definition of "$\subseteq$" at the end. If our goal is to prove "either $F \subseteq G$ or $F \subseteq H$", then we should conclude that one or the other of those claims holds: that *every* element of F is also an element of G, or else *every* element of F is also an element of H.

What we found was that every element of F is itself either an element of G or H, but we cannot decide collectively that *all* elements of F are elements of one or the other, G or H. Read through these last two paragraphs again to make sure you follow that logical observation. It might be easy to actually write up a "proof" for this claim and not realize that you've made a false step!

Identifying Errors

This recognition of an error is one of the skills we are developing, and it will be helpful in several ways. You'll notice that many exercises (some thus far, but many more as we move onwards) ask you to **find the flaw** in a proposed "proof" of some claim. By pointing out that there exists a flaw, we are perhaps helping you to find it (or them, as the case may be). Reading a proposed proof for logical, factual, and clarity errors is an essential skill. What's more, this careful reading of others' writing will necessarily make you a more critical reader of *your own* writing, and it will help you to catch potential errors like the one in the preceding paragraphs. Do not worry if you didn't catch it; now that you've seen it, you'll be on the lookout for similar mistakes in the future! Like we said, as well, this skill is ongoing development, and by the end of this book, you will be a great *reader* of mathematical proofs, as well as a great *writer*.

Counterexamples

So, now what do we do? We just recognized that our "proof" above did not work. Does this mean the claim is actually False? Actually, all this means (so far) is that our attempt at a proof failed. Maybe some other logical route will magically take us to the elusive conclusion.

Or, maybe the claim really is False. How could we show that? Think about the logical form of the claim: it says some statement holds true for *any* sets F, G, H. It says that the assumption $F \subset G \cup H$ will always imply, necessarily,

that $F \subseteq G$ or $F \subseteq H$. To show that this does *not* always happen, we need to find what's called a **counterexample**.

We will discuss all of these ideas again in the next chapter, when we formalize **logic**, but what you need to know for now is this: a **counterexample** is a specific, detailed, and described example that illustrates how a statement about "every ..." or "any ..." or "all possible ..." does *not* actually hold for every case. A counterexample amounts to **disproving** a statement that a whole class of objects has a certain property, by exhibiting one object in that class *without* that property.

Example

Let's see how the process of finding and stating a counterexample would work for our example above.

Example 3.9.10. Recall the claim:

For any sets F, G, H, if $F \subseteq G \cup H$, then either $F \subseteq G$ or $F \subseteq H$.

This claim is supposed to work for any sets F, G, H, so when we describe our counterexample, we better describe *exactly* what those three sets are going to be. We can't just explain our way around the issue and argue about how there might exist three sets out there with a certain property. Nope, we have to tell a reader exactly what they are by explicitly defining them. This is what the first line of our disproof of the claim will be, but we can't just jump right into that, because we don't know how to define them yet!

This is where the fun/work is: we need to play around with the desired properties of these sets to help us come up with an example. Recall that we want these sets to satisfy some properties: we should make sure the assumption $F \subseteq G \cup H$ holds True, but we want the conclusion—that either $F \subseteq G$ or $F \subseteq H$—to be False.

What does this mean? Well, we think you'll agree that, logically speaking, the "opposite" or "negation" of a statement like that would be "both $F \nsubseteq G$ and $F \nsubseteq H$". (This concept of **logical negation** will return in the next chapter; for now, we think you can understand it by applying the logical principles that guide your daily life. Soon, we will formalize this idea.)

We now have a specific goal: to find three sets F, G, H that satisfy all three of the following:

$$F \subseteq G \cup H$$
$$F \nsubseteq G$$
$$F \nsubseteq H$$

One thing left to consider is what "$\nsubseteq$" means. We have a definition of "$\subseteq$", so what is the "opposite" or "negation" of that? For $F \subseteq G$ to be true, we require that every element of F is also an element of G; so, if this *fails*, then we must have at least one element of F that is *not* an element of G. The same

observation applies to $F \not\subseteq H$. Now, we can restate our goals in a helpful way, by applying definitions:

> Every element of F is an element of either G or H
>
> There is at least one element of F that is not an element of G
>
> There is at least one element of F that is not an element of H

This will be incredibly helpful in finally finding our counterexample! We've boiled down all the essential parts of the claim and have restated the properties in a more intuitive way. The rest of the work is to just play around on some scratch paper and see what we can come up with. One approach is to draw a sort of "empty" Venn diagram, for F and G and H and their potential "overlaps", and then fill in enough elements so that the three above properties are satisfied.

The first condition requires the set F to "lie inside" both G and H, entirely; but, the second and third conditions require the existence of two elements of F, one of whom is not an element of G and the other of whom is not an element of H. That's all we need! A simple example, you might say, but an *effective* one, we say. Let's jump in and write up our disproof now:

Proof. The following claim is False:

> For any sets F, G, H, if $F \subseteq G \cup H$, then either $F \subseteq G$ or $F \subseteq H$.

We will disprove it with a counterexample.

Define $F = \{1, 2\}$ and $G = \{1\}$ and $H = \{2\}$.

Notice that $G \cup H = \{1, 2\}$. Since $F = G \cup H$, then certainly $F \subseteq G \cup H$. Thus, the hypothesis of the claim holds true.

However, notice that $2 \in F$ but $2 \notin G$. Thus, $F \not\subseteq G$.

Likewise, notice that $1 \in F$ but $1 \notin H$. Thus, $F \not\subseteq H$.

Therefore, the claim is False. □

One important lesson from this example is the following:

> Tour counterexample does not have to be the most interesting or complicated one, nor do you somehow need to characterize all possible counterexamples. We just need to see one counterexample, and we need to see how it works.

That's it! This is exactly what we did in the above proof: we defined all of the important objects (the three sets F, G, H), and then we pointed out and described all the relevant properties they had. We did not leave it to the reader to check that the counterexample works; we showed them the details. We did not argue that some such sets exist somewhere out there in the universe; we defined them explicitly.

This is important, and we expect your counterexamples to have similar proof structure to ours above. Most of the work will go on "behind the scenes", before

the proof starts, when you try to come up with your examples. Once you have it, though, just write it up much like we did.

3.9.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the definition of $\subseteq$? How do we use it to prove $A \subseteq B$?

(2) What does it mean for two sets to be equal?

(3) What is a *double-containment* proof?

(4) What is a *counterexample*?

(5) Suppose A, B, U are sets and $A, B \subseteq U$. Why can we prove $\overline{B} \subseteq \overline{A}$ to prove that $A \subseteq B$? Try to convince a friend that this is a valid technique.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) First, **prove** the following claim:

> For any sets A, B, C, the subset relationship $A - (B - C) \subseteq (A - B) \cup C$ holds.

Second, find a *counterexample* to the claim that those sets are actually always *equal*.

(2) Suppose A, B, C are sets and $A \subseteq B$. Prove that $A \times C \subseteq B \times C$.

(3) Suppose $A \subseteq C$ and $B \subseteq D$. Prove that $A \times B \subseteq C \times D$.

(4) Let $A = \{x \in \mathbb{R} \mid x^2 > 2x + 8\}$ and $B = \{x \in \mathbb{R} \mid x > 4\}$. For each of the following claims, either *prove* it is correct or provide a *counterexample* to show it is False.

(a) $A \subseteq B$

(b) $B \subseteq A$

(5) Let A, B, U be sets with $A, B \subseteq U$. Prove that $A - B = A \cap \overline{B}$ by a *double-containment argument*.

(6) Let $S = \{x \in \mathbb{R} \mid -2 < x < 5\}$ and $T = \{x \in \mathbb{R} \mid -4 \leq x \leq 3\}$. What is $S \cap \overline{T}$, in the context of $\mathbb{R}$? Identify a set and then *prove* it is correct, using a double-containment argument.

(7) **Prove** the following claim: If $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

(8) For every $n \in \mathbb{N}$ let $S_n = \left\{x \in \mathbb{R} \mid -\frac{1}{n} < x < \frac{1}{n}\right\}$. Prove that

$$\bigcap_{n \in \mathbb{N}} S_n = \{0\}$$

(9) Let $I = \{x \in \mathbb{R} \mid 0 < x < 1\}$. For every $x \in I$, define $S_x = \{y \in \mathbb{R} \mid x < y < x + 1\}$. Prove that

$$\bigcup_{x \in I} S_x = \{z \in \mathbb{R} \mid 0 < z < 2\}$$

(10) For every $n \in \mathbb{N}$, define the sets A_n and B_n by

$$A_n = \left\{x \in \mathbb{R} \mid 0 \leq x < \frac{n-1}{n}\right\}$$
$$B_n = \left\{y \in \mathbb{R} \mid -\frac{1}{n} < y < 1\right\}$$

Prove the following set equality by a double-containment argument:

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcap_{n \in \mathbb{N}} B_n$$

3.10 Summary

This was our first foray into some abstract conepts and results. We introduced the notion of a **set**, motivating it via several examples. We discussed the key relationships of being an **element** and being a **subset**, and pointed out how important is to distinguish the two! (Keeping the "Bag Analogy" in mind might help you in this regard.) We also discussed some notation, included *set-builder* notation. As we continue to move into more abstract mathematics, using correct, formal notation will be more important than ever to ensure that we are properly expressing our ideas. One key idea that came up is the notion of the *power set*, which represents a place where the *element* and *subset* relationships are both at work.

A discussion of set *operations* showed us how to combine sets and create new ones. All of these operations will be used throughout the remainder of our work in this book. We also showed how these operations can be *indexed*. This allows

us to use shorthand to write a union of several sets using just a few definitions and symbols. Again, these ideas will re-appear quite often throughout our work, so we will present many exercises relating to these ideas; we encourage you to attempt and work through as many as you can!

We saw a proof technique relating to sets: namely, **double-containment arguments**. This is a fundamental proof technique in mathematics. You will see us use it often, and you will find it appearing in other courses and studies, as well.

A couple of discussions came up that allowed us to touch upon some profound ideas in abstract set theory, although we couldn't completely dive into them. For one, *Russell's Paradox* showed us that there is no "set of all sets". For another, we talked about how the natural numbers can be formally defined in terms of *sets*. In practice, we won't use this definition, and will continue to rely on our intuitions about $\mathbb{N}$. However, we hope it was interesting and somehow informative to read such a discussion.

3.11 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them. Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ★.

Problem 3.11.1. ► For each of the following statements about elements and subsets, state whether it is True or False. Be prepared to defend your choice to a skeptical friend!

Throughout this problem, we will use the following definitions:

$$A = \{x \in \mathbb{Z} \mid -3 \leq x \leq 3\}$$
$$B = \{y \in \mathbb{Z} \mid -5 < y < 6\}$$
$$C = \left\{x \in \mathbb{R} \mid x^2 \geq 9\right\}$$
$$D = \{x \in \mathbb{R} \mid x < -3\}$$
$$E = \{n \in \mathbb{N} \mid n \text{ is even } \}$$

(a) $A \subseteq B$

(b) $C \cap D = \varnothing$

(c) $4 \in E \cap B$

(d) $\{4\} \subseteq A \cap E$

(e) $10 \in C - D$

(f) $A \cup B \supseteq C$

(g) $3 \in A \cap C$

(h) $0 \in (A - B) \cup D$

(i) $E \cap C \subseteq \mathbb{Z}$

(j) $0 \notin B - C$

Problem 3.11.2. ► Let $m, n \in \mathbb{N}$. Suppose $m \leq n$. Explain why $\mathcal{P}([m]) \subseteq \mathcal{P}([n])$.

Problem 3.11.3. Look back at Problem 7 in Section 3.9. We proved that whenever two sets satisfy $A \subseteq B$, then they must also satisfy $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. Read through that proof, too, to remind yourself of the details.

Now, does this claim "work the other way"? That is, suppose $\mathcal{P}(A) \subseteq \mathcal{P}(B)$. Can you prove that $A \subseteq B$ is also true? Or can you find an example where this is not true?

Problem 3.11.4. Rewrite the following sentences using the "set-builder notation" to define a set. Then, if possible, **write out** all the elements of the set, using set braces; if not possible, explain why not and write out **three** example elements of the set.

(a) Let A be the set of all natural numbers whose squares are less than 39.

(b) Let B be the set of all real numbers that are roots of the equation $x^2 - 3x - 10 = 0$.

(c) Let C be the set of pairs of integers whose sum is non-negative.

(d) Let D be the set of pairs of real numbers whose first coordinate is positive and whose second coordinate is negative and whose sum is positive.

Problem 3.11.5. Define the following sets:

$$A = \left\{x \in \mathbb{R} \mid x^2 - x - 12 > 0\right\}$$
$$B = \{y \in \mathbb{R} \mid -3 < y < 4\}$$

Prove that $A = B$.

Problem 3.11.6. Let X be the set of students at your school.

Identify a property $P(x)$ such that $A := \{x \in X \mid P(x)\}$ is a proper subset of X and $A \neq \varnothing$.

Then, identify a property $Q(x)$ such that $B := \{x \in X \mid Q(x)\}$ is a proper subset of A (i.e. $B \subset A$) and $B \neq \varnothing$.

Problem 3.11.7. Let A, B, and C be sets with $A \subseteq C$ and $B \subseteq C$.

(a) Draw a Venn diagram for the sets $\overline{A} \cap \overline{B}$ and $\overline{(A \cap B)}$.

(b) Prove that $\overline{A} \cap \overline{B} \subseteq \overline{(A \cap B)}$.

(c) Define specific sets A, B, C such that the containment is *strict*, i.e. $\overline{A} \cap \overline{B} \subset \overline{(A \cap B)}$

(d) Define specific sets A, B, C such that $\overline{A} \cap \overline{B} = \overline{(A \cap B)}$.

Problem 3.11.8. Let $S = \left\{(m, n) \in \mathbb{Z} \times \mathbb{Z} \mid m = n^2\right\}$. How does S compare to the set $T = \{(m, n) \in \mathbb{Z} \times \mathbb{Z} \mid n = m^2\}$? If one is a subset of the other, prove it. If not, provide examples to show this.

Problem 3.11.9. Let (a, b) be a point on the Cartesian plane, i.e. $(a, b) \in \mathbb{R} \times \mathbb{R}$. Let ε (the Greek letter *epsilon*) be a nonnegative real number, i.e. $\varepsilon \in \mathbb{R}$ and $\varepsilon \geq 0$.

Let $C_{(a,b),\varepsilon}$ be the set of real numbers that are "close" to (a, b), defined as follows:

$$C_{(a,b),\varepsilon} = \left\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid \sqrt{(x-a)^2 + (y-b)^2} < \varepsilon\right\}$$

1. Come up with a geometric description of the set $C_{(a,b),\varepsilon}$.

   What happens to the set as we change a and b?

   What happens as we change ε?

2. What is $C_{(0,0),1} \cap C_{(0,0),2}$?

3. What is $C_{(0,0),1} \cup C_{(0,0),2}$?

4. What is $C_{(0,0),1} \cap C_{(2,2),1}$?

Problem 3.11.10. Consider the (false!) claim that

$$\bigcup_{n \in \mathbb{N}} \mathcal{P}([n]) = \mathcal{P}(\mathbb{N})$$

(a) What is wrong with the following "proof" of the claim? Point out any error(s) and explain why it/they ruin the "proof".

> First, we will show that
>
> $$\bigcup_{n \in \mathbb{N}} \mathcal{P}([n]) \subseteq \mathcal{P}(\mathbb{N})$$

Consider an arbitrary element X of the union on the left.

By the definition of an indexed union, we know there exists some $k \in \mathbb{N}$ such that $X \subseteq [k]$.

Since $[k] \subseteq \mathbb{N}$, and $X \subseteq [k]$, we deduce that $X \subseteq \mathbb{N}$.

Thus, $X \in \mathcal{P}(\mathbb{N})$.

Second, we will prove the "$\subseteq$" relationship holds in the other direction, as well.

Consider an arbitrary $Y \subseteq \mathbb{N}$.

By the definition of subset, and the fact that Y is a set of natural numbers, we know there exists some $\ell \in \mathbb{N}$ such that $Y \subseteq [\ell]$.

By the definition of indexed union, we know $Y \in \bigcup_{n \in \mathbb{N}} \mathcal{P}([n])$.

Since we have shown $\subseteq$ and $\supseteq$, we know the two sets are equal.

(b) Disprove the claim by defining an **explicit** example of a set S such that

$$S \in \mathcal{P}(\mathbb{N}) \qquad \text{and} \qquad S \notin \bigcup_{n \in \mathbb{N}} \mathcal{P}([n])$$

Problem 3.11.11. Let $A = [3] \times [4]$. (Remember that $[n] = \{1, 2, 3, \ldots, n\}$.)

Let $B = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid 0 \leq 3x - y + 1 \leq 9\}$.

(a) **Prove** that $A \subseteq B$.

(b) Is it true that $A = B$? Why or why not? **Prove** your claim.

Problem 3.11.12. Let $n \in \mathbb{N}$ be a fixed natural number. Let $S = [n] \times [n]$. Let T be the set

$$T = \left\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid 0 \leq nx + y - (n + 1) \leq n^2 - 1\right\}$$

Prove that $S \subseteq T$ but $S \neq T$.

Problem 3.11.13. Suppose A and B are sets.

(a) **Prove** that

$$\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$$

(b) Provide an **explicit** example of A and B where the containment in (a) is **strict**.

Problem 3.11.14. Let S and T be sets whose elements are sets, themselves. Suppose that $S \subseteq T$.

Prove that

$$\bigcup_{X \in S} X \subseteq \bigcup_{Y \in T} Y$$

Problem 3.11.15. Let A, B, C, D be sets.

(a) **Prove** that
$$(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$$

(b) Provide an **explicit** example of A, B, C, D where the containment in (a) is **strict**.

Problem 3.11.16. Let A, B, C be sets. Prove that
$$A \times (B \cap C) = (A \times B) \cap (A \times C)$$
and
$$A \times (B - C) = (A \times B) - (A \times C)$$

Problem 3.11.17. Let X, Y, Z be sets. Prove that $(X \cup Y) - Z \subseteq X \cup (Y - Z)$ but equality *need not* hold.

Problem 3.11.18. Find an example of a set S such that $S \in \mathcal{P}(\mathbb{N})$ and S contains exactly 4 elements.

Then, find an example of a set T such that $T \subseteq \mathcal{P}(\mathbb{N})$ and T contains exactly 4 elements.

Problem 3.11.19. Find examples of sets R, S, T such that $R \in S$ and $S \in T$ and $R \subseteq T$ but $R \notin T$.

Problem 3.11.20. Identify what each of the following sets are, and **prove** your claims.
$$\bigcap_{n \in \mathbb{N}} [n] \qquad \text{and} \qquad \bigcup_{n \in \mathbb{N}} [n]$$

Problem 3.11.21. Let $I = \{-1, 0, 1\}$. For each $i \in I$, define $A_i = \{i - 2, i - 1, i, i + 1, i + 2\}$ and $B_i = \{-2i, -i, i, 2i\}$.

(a) Write out the elements of $\bigcup_{i \in I} A_i$.

(b) Write out the elements of $\bigcap_{i \in I} A_i$.

(c) Write out the elements of $\bigcup_{i \in I} B_i$.

(d) Write out the elements of $\bigcap_{i \in I} B_i$.

(e) Use your answers above to write out the elements of $\left(\bigcup_{i \in I} A_i\right) - \left(\bigcup_{i \in I} B_i\right)$.

(f) Use your answers above to write out the elements of $\left(\bigcap_{i \in I} A_i\right) - \left(\bigcap_{i \in I} B_i\right)$.

(g) Write out the elements of $\bigcup_{i \in I}(A_i - B_i)$. How does this compare to your answer in (e)?

(h) Write out the elements of $\bigcap_{i \in I}(A_i - B_i)$. How does this compare to your answer in (f)?

Problem 3.11.22. In this problem, we are going to "prove" the existence of the negative integers! We say "prove" because we won't really understand what we've done until later but, trust us, it's what we're doing.

Because of this goal, you cannot **assume** any integers strictly less than 0 exist, so your algebraic steps, especially in part (d), should not involve any terms that might be negative.

That is, if you consider an equation like

$$x + y = x + z$$

we **can** deduce that $y = z$, by subtracting x from both sides, since $x - x = 0$.

However, if we consider an equation like

$$x + y = z + w$$

we **cannot** deduce that $x - z = w - y$. Perhaps $y > w$, so $w - y$ does not exist in our context ...

Let $P = \mathbb{N} \times \mathbb{N}$. Define the set R by

$$R = \{((a, b), (c, d)) \in P \times P \mid a + d = b + c\}$$

(a) Find three different pairs (c, d) such that $((1, 4), (c, d)) \in R$.

(b) Let $(a, b) \in P$. Prove that $((a, b), (a, b)) \in R$.

(c) Let $((a, b), (c, d)) \in R$. Prove that $((c, d), (a, b)) \in R$, as well.

(d) Assume $((a, b), (c, d)) \in R$ and $((c, d), (e, f)) \in R$.
Prove that $((a, b), (e, f)) \in R$, as well.

3.12 Lookahead

Now that we've introduced sets, defined them, seen many examples, and talked about operations and how to manipulate sets, it's time to move on to logic. We've already previewed some important logical ideas, specifically in Section 3.9 on how to write **proofs** about sets. In the next Chapter, we will make all of these logical ideas more formal, explicit and rigorous. We will develop some

notation and grammar that will help us express logical ideas more precisely and concisely. We will use these to express our mathematical thoughts in a common language and communicate our ideas with others. In short, we will be able to confidently talk and write about mathematics!

Chapter 4

Logic: The Mathematical Language

4.1 Introduction

We are moving on to learn about the language of mathematics! We will learn how to express our ideas formally and precisely and concisely. This will require learning some new terminology and notation, all the while thinking and writing in a more formal way. Ultimately, this will allow us to solve problems and write good, clear, and correct mathematical **proofs**.

4.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to ...

- Define variable propositions using proper notation.

- Define mathematical statements using quantifiers and other proper notation, and characterize sentences that are not proper mathematical statements.
- Understand and explain the difference between two types of quantifiers, as well as how they are used.
- Define and understand the meanings of several logical connectives, as well as use them to define more complicated mathematical statements.
- Apply proof techniques to create formal arguments that demonstrate the truth of a mathematical statement.
- Compare and contrast different types of proof techniques, as well as understand when and how to use them, depending on the situation.

4.1.2 Segue from previous chapter

We introduced sets to have some standard, fundamental mathematical objects to work with. You probably noticed, though, that we used a lot of phrases like "If ..., then ..." and "for *every*" and "there *exists*" and "for *at least one*" and "and" and "or" and True and False and so on and so forth ... We relied on your intuitive understanding, and previous knowledge, of these concepts. As a living, breathing human being who converses with others, you have some kind of understanding of what **logic** is. Our goal now is to build upon those intuitions, and help you learn to read and write and speak mathematics.

4.1.3 Motivation

In mathematics, we are interested in identifying True claims and subsequently explaining to others *how* and *why* we know those claims are True. Thus far, we have already presumed some familiarity with logical terminology and truth. For instance, look back at the assumptions of the PMI (Principle of Mathematical Induction, Theorem 3.8.5). We needed to know that *if* $P(k)$ is true *then* $P(k+1)$ is true. What does this mean? What does this say about how the statements $P(k)$ and $P(k+1)$ are connected? What does it even mean for something to be True?!?!

Our goals for this section are many, but the major emphasis is on defining and identifying what types of statements in mathematics are meaningful and interesting. Once we do that, we can figure out how to express those statements in concise and precise terms. Ultimately, we will learn how to apply general techniques to **prove** that those statements are True (or False, as the case may be).

4.1.4 Goals and Warnings for the Reader

Keep this in mind throughout this chapter:

You are learning a **language**.

Some of this material will seem difficult, some will seem boring, and some might seem both! But this is all essential.

Have you ever learned another language? Think back to a foreign language class you took in school, perhaps. How did you start? We bet that you didn't jump right in and try to write beautiful poetry. You learned the basic grammar and syntax and vocabulary. You learned important articles, like "the" and "an". You learned basic verbs like "to be" and "to have" and how to conjugate them. You learned some common nouns, like "apple" and "dog" and "friend". From there, you started to piece phrases together and, over time, you learned to create more complex sentences from all of the tools you developed. All along, you probably had some great ideas for wonderful sentences in mind, but you just didn't know how to express them in your new language until you learned the necessary words and grammar.

We will be doing exactly this in the current chapter, but here our language is **mathematics**. You might have some great ideas for wonderful mathematical sentences in mind, but you just aren't sure how to express them. Interestingly enough, as well, we've already been "speaking" a lot of mathematics with each other! We've solved some puzzles, we introduced a proof technique (induction), and we worked with some building blocks of mathematical objects (sets); all along, we've made sure that we've understood each other, being verbose with our writing and explaining lots of details. In a way, we've been communicating without setting down a common language. This is a lot like how you'd survive in a foreign country without knowing the language: you'd use a lot of hand gestures and charades, you'd try to listen to others speak and pick up some key words, you'd draw pictures and make noises, and so on. This is all well and good if you're just on a week-long vacation, say, but if you're going to *live* there, you'll have a lot more work ahead of you.

This is precisely the situation we face now. We'd like to inhabit this world of mathematics, so we need to settle down and really learn the language of its people. Once we get through that, we'll feel more at home, like native speakers. Then, we can maybe be a little less careful with our words and grammar, use some slang or abbreviations or common idioms. (Think of some English examples of phrases and sentences that technically make no sense, grammatically or in terms of vocabulary, but are still understood by your fellow speakers.) But only then can we do so.

In the meantime, we will be much more formal and pedantic with our language. *If we don't force ourselves to do this now, we won't truly all speak the same mathematical language.*

4.2 Mathematical Statements

Our first step is to discuss what types of sentences are even reasonable to consider as mathematical truths that need to be proven or disproven. Completing this step is actually quite difficult! Many authors tend to gloss over this subject

or offer a simple definition that ignores the many subtleties of mathematical language and logic. We feel tied, as well, because the time and space provided in this book/course are not sufficient to properly study the field of abstract logical theory. We encourage you to investigate some books or websites that contain relevant information. For the current context, we will have to sweep many details under the rug, so to speak. Suffice it to say, though, there is a very deep, rich, and fruitful field of mathematical research concerning exactly what we will be discussing here in a more heuristic way.

Remember that we mentioned we will have to assume the existence of the real numbers $\mathbb{R}$ and their usual arithmetic properties. Likewise, we will assume many of the results and concepts of mathematical logic, often without even realizing it (until we point it out for you). These details can be studied more in-depth later on in your mathematical careers.

4.2.1 Definition

For now, let us discuss what we mean by a **mathematical statement**. We want this term to encapsulate the kinds of "things" that we can prove or disprove.

Mathematics is unique among the sciences in that the results of this field are **proven** rigorously, and not hypothesized and then "confirmed" via laboratory experiments or real-world observations. In mathematics, we assume a set of common **axioms** and then follow rigorous logical inferences to deduce truths from these axioms (and from other truths we have proven thus far). If we encounter a falsity, we would have to show or demonstrate that it is, indeed, False.

With these ideas in mind, we can now consider and agree on several examples of what such a **mathematical statement** or **proposition** could be. (We have even proven a few already!) For instance, the sentence

> For any real numbers $x, y \in \mathbb{R}$, the inequality $2xy \leq x^2 + y^2$ holds.

is a valid mathematical statement. In fact, it is True, and we will prove it later on in Section 4.9.2. (It is sometimes referred to as the *AGM Inequality*, short for the *Arithmetic-Geometric Mean Inequality*.) We should point out that the word "holds" is often used in mathematics to mean "holds true" or "is a true statement".

Here's another example of a mathematical statement:

> For any sets S, T, U, if $S \cap T \subseteq U$ then $S \subseteq U$ or $T \subseteq U$.

This statement, however, is False, as the following **counterexample** demonstrates:

> Let $S = \{1, 2, 3\}$ and $T = \{2, 3, 4\}$ and $U = \{2, 3, 5\}$.
>
> Observe that $S \cap T = \{2, 3\} \subseteq U$ but $S \not\subseteq U$ and $T \not\subseteq U$.

Why does this example *disprove* the statement? Do you understand? Can you explain it? We will discuss that in more detail later on in this chapter, but we

hope that, for now, we all somehow recognize that this example accomplishes exactly that.

We can also agree that a sentence like

$$\text{Why do we have class at 9:00 am?!}$$

is definitely *not* a mathematical statement. It's a perfectly valid English sentence, but it is not meaningful, mathematically speaking: we can't *prove* or *disprove* it.

Likewise, the sentence

$$x^2 - 1 = 0$$

is *not* a mathematical statement, despite being composed entirely of mathematical symbols. The problem is that we cannot verify whether it is True or False purely from axioms and logical inferences. This statement *depends* on x, whatever that value is (i.e. x is a **variable**) and without imposing extra assumptions about it, we cannot declare whether this sentence is True or False. This type of sentence will be referred to later as a **variable proposition**: its truth depends on a variable inside the sentence.

All of these observations and examples/non-examples motivate the following definition:

Definition 4.2.1. *A* **mathematical statement** *(or* **proposition** *or* **logical statement***) is a grammatically correct sentence (or string of sentences), composed of English words/symbols and mathematical symbols that has* exactly *one truth value, either* True *or* False.

4.2.2 Examples and Non-examples

By "grammatically-correct" we mean that the words and symbols contained in the sentence are used and combined correctly and make sense. This eliminates strings of symbols/words that are nonsensical when placed together, like

$$1+=2 \quad , \quad \text{Brendan}^2 = 1 \quad , \quad \{\{\varnothing\}\} - 7 > 5\pi \quad , \quad \text{You am smart}$$

For instance, the third one above is not a mathematical statement because $\{\{\varnothing\}\}$ is not a number, so we don't know how to interpret "subtracting 7" from that set.

By "has *exactly* one truth value", we mean that the statement should be either True or False, but certainly cannot be both or neither or something else in between. This eliminates the "$x^2 - 1 = 0$" example above, because it has no truth value. (Without a declaration of what x is, we cannot decide, either way.)

Not Knowing the Truth Value

One strange/interesting/complicated aspect of our definition is that we might not know the truth value for a given statement, even though we can be sure that there is *only* one such value. By way of illustration, consider the following statement:

> Any even natural number greater than or equal to 4 can be written as the sum of two prime numbers.

Is this statement True or False? If you have a proof or disproof, then the world of mathematics would love to see it! The statement above is known as the **Goldbach Conjecture** and it is a very famous unsolved (for now, we hope!) problem in mathematics. Nobody knows yet whether the claim is True or False, but it is certainly the case that *only one* of those truth values applies. That is, this statement cannot be both True and False, nor can it somehow be somewhere in between. Either all even natural numbers greater than or equal to 4 *do* have the stated property, or there is at least one that does *not* have the property. We can state this "either/or" property even without yet knowing which of the two possibilities is the correct one. As such, this sentence *does* actually satisfy our definition of *mathematical statement*.

(Terminology note: In general, a **conjecture** is a claim that someone believes to be true but has not yet been proven/disproven.)

Paradoxical Sentences

One way to have a sentence that does *not* have a truth value is to create a **paradox**. Consider this sentence:

> This sentence is False.

Pretty weird, right? The sentence itself is asserting something about its own truth value. Let's try to analyze what truth value it has:

- Let's say the sentence is True. Then, the sentence itself tells us that it is, in fact, False.
- Let's say the sentence is False. Similarly, then, the sentence tells us that it is, in fact, True.

This cannot work! This sentence is somehow both True and False at the same time, or somehow neither, or ... Whatever it is, it's a bad idea. We do *not* want to deal with strangeness like this in mathematics, so our definition disallows this sentence as a mathematical statement.

(*Question:* What happens if you *do* allow sentences like this to be proper mathematical statements? What if you don't adhere to the principle that every sentence we care about must be either True or False? Think about it! Is this somehow *wrong*, or is it just a different mathematical universe? ...)

In general, **self-referential** sentences like the one above (that is, sentences that make reference to themselves) are quite bizarre and can produce some paradoxes that we want to disallow.

A variant of the above paradoxical claim is given in a cartoon drawing, wherein Pinocchio says, "My nose will grow now!" Does it then grow? If he's telling the truth, then it will grow, but that only happens when he's lying! If

he's lying, then his nose will grow (by definition), but then his statement is actually true! Yikes!

Source: http://www.the-drone.com/magazine/wp-content/uploads/2010/04/BLA6.jpg

An even stranger example of this phenomenon is *Quine's Paradox*:

> "Yields falsehood when preceded by its quotation" yields falsehood when preceded by its quotation.

We'll let you think about that one on your own. Suffice it to say that paradoxical claims like this are too ill-behaved for us to worry about. This is why our definition outlaws them.

4.2.3 Variable Propositions

Other examples of sentences that are not mathematical statements are sentences that involve **unquantified variables**. For instance, take the sentence

> " $x^2 - 1 = 0$ "

This is certainly grammatically correct and we can make sense of it, but what is its truth value? We don't know! If $x = 1$, then the sentence is True, but if $x = 8$, it is False, and if $x = \mathbb{N}$ or $x =$ Brendan, then the sentence doesn't even make sense! As such, we want to disallow sentences like this, as well. These types of sentences are useful and common, though; we will call them **variable propositions** because they make a claim that *depends* on some variable.

In the case of the above sentence, we might define $P(x)$ to be the variable proposition "$x^2 - 1 = 0$". We would usually write this declaration as

> Let $P(x)$ be the statement "$x^2 - 1 = 0$".

It is common to use capital letters to denote variable propositions and mathematical statements, and lowercase letters to denote the variables contained therein. (This is not a *requirement*, though, merely a common convention.)

With this variable proposition now defined, we can create proper mathematical statements by *assigning* particular values to the variable x in the expression. We can say that $P(1)$ is True and $P(0)$ is False. We can also make **quantified** claims about $P(x)$. For instance, we claim that the following sentence is a True mathematical statement:

> There exists an $x \in \mathbb{R}$ such that the proposition $P(x)$ is True.

whereas the following sentence is a False mathematical statement:

> For every $x \in \mathbb{R}$, the proposition $P(x)$ is True.

Think about why these statements have the truth values we have claimed. Can you see why they are mathematical statements to begin with? How would you *prove* these claims?

Defining Variable Propositions

Notice the format we used to define variable propositions, like the one above: (1) We give the proposition a letter name (like P); (2) we indicate its dependence on some number of variables, each of which has a letter (like x and y); (3) we put quotes around the actual proposition itself; and (4) we don't include any new letters that have no meaning within the context of the proposition.

This format has been chosen carefully because it is precise and unambiguous. It assigns a meaning to every letter in the proposition and clearly distinguishes what is and isn't part of the proposition.

For example, the following are **BAD** "definitions" of variable propositions. We will give you some reasons as to why they are bad and provide some proper amendments of the propositions.

- Let $Q(y)$ be the proposition "$x < 0$".

  Reason: What is x? Where is y? We have *no idea* what x is, inside the context of the proposition, so this is a poor definition.

  If we had said

  > Let $Q(x)$ be the statement "$x < 0$".

  that would have been perfect. The variable inside the parentheses is the one that appears in the statement in quotes later. Great.

- Let $P(x)$ be the proposition $x^2 \geq 0$, for every $x \in \mathbb{R}$.

  Reason: Does the writer of this sentence want to assert that $x^2 \geq 0$, no matter what $x \in \mathbb{R}$ is? Is that phrase "for every $x \in \mathbb{R}$" meant to be *part* of the proposition, or not?

  If we interpret this to mean that $P(x)$ is defined as "$x^2 \geq 0$", and this definition is made for every $x \in \mathbb{R}$, then ... okay, that might be reasonable.

  However, if we interpret this to mean that $P(x)$ is defined as "$x^2 \geq 0$ for every $x \in \mathbb{R}$" then ... well, this is certainly *different*. In fact, it's not even

a properly-defined proposition! The proposition $P(x)$ should *depend* on the input value x, but it shouldn't be allowed to *change* or further *quantify* that variable *inside* the proposition!

The way this proposition was originally written, there are two possible interpretations and they are very different. Accordingly, this is a poor definition.

If we had said,

> Let $P(x)$ be the statement "$x^2 \geq 0$", defined for every $x \in \mathbb{R}$.

that would have been fine. As we mention below, as well, we don't technically have to tell the reader which values of x we want the proposition defined for. Perhaps this is just some helpful information to include, though, so it doesn't hurt.

- Let $T(x, y) =$ "$x^2 - 7 = y$".

  Reason: Ugh! What does "=" mean in this context? That symbol applies when we wish to compare two *numbers* and say they are *equal* in value (or two *sets* and say they are equal in terms of their elements). The object $T(x, y)$ is meant to be a *mathematical statement*, something that is either True or False. Thus, it does not have a numerical value to compare with anything else.

  Likewise, given a value for x and y, the statement "$x^2 - 7 = y$" is either True or False, so it makes no sense to say that equation "equals" something else. It has a truth value, not a numerical value.

  If we had said,

  > Let $T(x, y)$ be "$x^2 - 7 = y$".

  that would have been perfect.

Okay, that's enough *non*-examples for now. We don't want to put any bad ideas in your head, really! However, from past experience, we know that these are common ways for students to write propositions (either accidentally, or without realizing *why* they're wrong), so we felt compelled to share.

One final note on variable propositions. It is not essential to say where the variables come from when defining a proposition. That can be filled in later when the proposition is invoked, or when a specific value of a variable is used or quantified. That is, we can make the definition

> Let $T(x, y)$ be "$x^2 - 7 = y$".

without needing to specify whether x and y are natural numbers, or integers, or complex numbers, or anything like that. Later on, we can say that $T(3, 2)$ is True and $T(\pi, -1)$ is False, and that $T(\varnothing, \mathbb{N})$ has *no meaning*, but we wouldn't need to somehow anticipate any of those interpretations when defining $T(x, y)$.

4.2.4 Word Order Matters!

The notion of **quantifying variables** is something we will discuss in detail in the very next section. For now, we want to consider one more striking example of a mathematical statement that illustrates the importance of **word order** in sentences. Analyzing the structure of sentences like the following will be a major goal of the following section, as well.

There is a real number y such that $y = x^3$ for every real number x.

What does this claim say? It says we can find a number $y \in \mathbb{R}$ such that $y = x^3$ is True, *no matter what* $x \in \mathbb{R}$ is. This is ridiculous! How can there be a single number that is the cube of *all* numbers? This sentence is, indeed, a mathematical statement, but it is decidedly False. But what about the following claim?

For every real number x there is a real number y such that $y = x^3$.

This one is True! Do you see the difference between the two sentences? They contain exactly the same words and symbols, but in a different order. Whereas the former sentence asserts that there is some number that is the cube of every real number (which is False), the latter asserts that every real number has a cube root, which is True. This example emphasizes the importance of word order.

4.2.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What are the important, defining properties of a mathematical statement?

(2) What is the difference between a mathematical statement and a variable proposition?

(3) Why is the Goldbach Conjecture a mathematical statement?

(4) What is wrong with the following attempt at defining a variable proposition?

Let $Q(x, y, z)$ be $7x - 5y + z$

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) For each of the following sentences, decide whether it is a **mathematical statement** or not. If it is, decide whether it is True or False. If it is not, explain why.

   (a) $142857 \cdot 5 = 714285$

   (b) For every $n \in \mathbb{N}$, $\displaystyle\sum_{k\in[n]} k = \frac{n(n+1)}{2}$.

   (c) For any sets A and B, if $A \subseteq B$, then $B \subseteq A$.

   (d) For any sets A and B, if $A \subseteq B$, then $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

   (e) Math is cool.

   (f) $1 + 2 = 0$

   (g) For any $x, y \in \mathbb{Z}$, if $x \cdot y$ is even, then x and y are both even.

   (h) For any $x, y \in \mathbb{Z}$, if x and y are both even, then $x \cdot y$ is even.

   (i) $1+ = 2$

   (j) $-5 + \mathbb{Z} \geq \pi$

   (k) $x = 7$

   (l) This sentence is not True.

(2) Look back through the first three chapters and identify some examples and non-examples of mathematical statements.

   Can you also find any variable propositions? Are they written in the way we specified in this section? Can you amend them so that they are written properly?

(3) Write a proper definition of a variable proposition that is true when two inputted values have a non-negative sum.

   Then, find two instances each of when the proposition is True and when it is False.

(4) Let S be the set $\{1, 2, 3, 6, 8, 10\}$.

   (a) Write a proper definition of a variable proposition that inputs two variables and decides whether the absolute value of their difference is an element of S. Then, find two instances each of when the proposition is True and when it is False.

   (b) Write a proper definition of a variable proposition that inputs two sets and determines whether their intersection is a subet of S. Then, find two instances each of when the proposition is True and when it is False.

   (Note: Given any set X and any object x, it must be that either $x \in X$ or $x \notin X$.)

(5) Come up with another mathematical statement that is True but becomes False when we switch the order of some words. (See the example in the last subsection for some inspiration.)

(6) For each of the following attempts at defining a variable proposition, determine if it is correct or not. Note: This does not mean determine if it is True or False; rather, we want to know whether the statement is well-written and sensible.

If an attempt is incorrect for some reason, explain that reason and write a new statement that fixes that error.

(a) Let $P(x)$ be "$x > 1$".

(b) Let $Q(x)$ be the proposition "$x^2 - 1 > 0$".

(c) Let $R(a, b)$ be "$a^3 = b$", for every $a, b \in \mathbb{R}$.

(d) Let $P(x)$ be $x > 1$.

(e) Let $T(z) =$ "z is prime".

(f) Let $Q(x)$ be the proposition "$x^2 - 1 > 0$", for every $x \in \mathbb{R}$.

(g) For every $x \in \mathbb{R}$, let $Q(x)$ be "$x^2 - 1 > 0$".

(h) Let $S(a)$ be "$b^2 > 4$".

(i) Let $Q(x)$ be $x^2 - 1 > 0$ for every $x \in \mathbb{R}$.

4.3 Quantifiers: Existential and Universal

We will now introduce some convenient notation that allows us to shorten some statements we have seen so far and express wordy, language-based phrases with mathematical symbols. Another benefit of the forthcoming notation is that we will be able to more easily express and analyze mathematical statements. Specifically, we will now introduce the symbols "$\forall$" and "$\exists$".

Definition 4.3.1. *The symbol "$\forall$" stands for the phrase "**for all**".*

*The symbol "$\exists$" stands for the phrase "**there exists**".*

We call "$\forall$" the universal quantifier *and "$\exists$" the* existential quantifier*.*

A mathematical statement beginning with "$\forall$" is said to be "universally quantified", and one beginning with "$\exists$" is said to be "existentially quantified".

4.3.1 Usage and notation

Other common phrases that "$\forall$" replaces are "for every" and "for arbitrary" or "whenever" and "given any" and even "if".

Other common phrases that "$\exists$" replaces are "for some" and "there is at least one" and "there is" and even "some".

Example 4.3.2. Let's consider some simple examples first, to get our feet wet. In each case, we are looking to express a mathematical thought using these symbols, or trying to interpret a quantified statement in a more "wordy" way.

- "Every real number squared is non-negative."

  This is a straightforward statement that is, in fact, True. We would write it as:

  $$\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0$$

  The "big dot" separates the quantified part of the statement from the claim made about the variable x (which was introduced in the quantification).

  Another way to write this would be:

  > Define $S(x)$ to be "$x^2 \geq 0$". Then the claim is: $\forall x \in \mathbb{R}\text{.}\ S(x)$.

- "There is a subset of $\mathbb{N}$ that has 7 as an element."

  This is an *existence* claim. It asserts that we can find an object with a particular property. We would write it as:

  $$\exists S \in \mathcal{P}(\mathbb{N})\text{.}\ 7 \in S$$

  Remember that $\mathcal{P}(\mathbb{N})$ is the *power set* of $\mathbb{N}$, the set of all subsets of $\mathbb{N}$; thus, saying $S \in \mathcal{P}(\mathbb{N})$ means $S \subseteq \mathbb{N}$, as desired.

- "Every integer has an *additive inverse* (i.e. a number that, when added to the original number, yields 0)."

  This idea of an "additive inverse" is a general concept that applies to some mathematical objects known as *rings* and *fields*. We won't discuss those objects in this book, but you will touch on them in any course on abstract algebra.

  We would write this claim as

  $$\forall a \in \mathbb{Z}\text{.}\ \exists b \in \mathbb{Z}\text{.}\ a + b = 0$$

  and we would read this aloud as

  > For any integer a, there exists an integer b such that $a + b = 0$.

  or perhaps

  > No matter what integer a we are given, we can find an integer b with the property that $a + b = 0$.

  Again, we could shorten the notation slightly by defining $I(a, b)$ to be "$a + b = 0$", and then writing the claim as

  $$\forall a \in \mathbb{Z}\text{.}\ \exists b \in \mathbb{Z}\text{.}\ I(a, b)$$

Example 4.3.3. Here are some examples of proper usage of "$\forall$", and some equivalent formulations of how to use this symbol.

- $\forall x \in \mathbb{R}\,\boldsymbol{.}\; x^2 \geq 0$
- For all real numbers x, we have $x^2 \geq 0$.
- Every real number x satisfies $x^2 \geq 0$.
- Whenever x is a real number, we know $x^2 \geq 0$.

Likewise, here are some examples of proper usage, and equivalent formulations, of the symbol "$\exists$".

- $\exists x \in \mathbb{R}\,\boldsymbol{.}\; x^2 - 4x + 4 = 0$
- There exists a real number x such that $x^2 - 4x + 4 = 0$.
- There is a real number x that satisfies $x^2 - 4x + 4 = 0$.
- Some real number x has the property taht $x^2 - 4x + 4 = 0$.

Reading Quantified Statements Aloud

Example 4.3.4. Now for some harder examples. Let's look back at the phrases we wrote at the end of the last section and express them using this new notation. Consider this statement:

> There is a real number y such that $y = x^3$ for every real number x.

To express thi in symbolic form, we will define $P(x, y)$ to be the proposition "$y = x^3$" and then write the statement as

$$\exists y \in \mathbb{R}\,\boldsymbol{.}\; \forall x \in \mathbb{R}\,\boldsymbol{.}\; P(x, y)$$

This is correct, logically-speaking, but it is rather terse. For now, we will sometimes rewrite the statement using some "helping words" to aid our reading of the statement. In particular, we would say such words when reading the sentence aloud, so by occasionally writing them here, we provide you with some extra practice in interpreting logical notation verbally. We would read the above statement aloud as

> There exists a real number y such that, for every real number x, the statement $P(x, y)$ holds.

The phrase "*such that*" is a "helper phrase" that links an existential quantification to the rest of the phrase. The next subsection contains some important information about when and how to use this helper phrase!

The "big dot" between the quantified parts of the statement above just serves to separate the pieces of the statement and make it easier to read. It corresponds to a pause or rest in speaking, like a comma, but sometimes it has a vocalized meaning (like the "such that" after the "$\exists y \in \mathbb{R}$" part).

We don't want to use commas, though, because we already use them for other meanings. For example, we write

$$x, y \in S$$

to mean "both x and y are elements of the set S". The "big dot" is just a different symbol to use.

Since our mathematical careers are still young, relatively speaking, we encourage you to sometimes write the helping phrases like "such that" and "holds True" to guide your understanding, whenever possible. This reminds you what the sentences mean and helps you practice reading and writing statements like this in such a condensed form. Remember that you are learning a *language* here and you need to practice *translating* sentences from one language you know (English) to another (mathematics). For instance, you might want to write out the line above as

$$\exists y \in \mathbb{R} \text{ such that } \forall x \in \mathbb{R}\text{.}\ P(x, y) \text{ is True.}$$

or, at least, say it this way in your head.

(By the way, when writing on a white/chalkboard or on paper, it's common to write "s.t." in place of "such that", to save a few moments of writing. That just goes to show how ubiquitous the phrase "such that" is in mathematical writing; we already have an agreed-upon abbreviation for it!)

4.3.2 The phrase "such that", and the order of quantifiers

Notice that the helping phrase "such that" always follows an *existential* quantification, and *only* such a quantification. This is because a claim with "$\exists$" asserts something about the existence of an object with a certain property, and the rest of the sentence is the description of that special property. Thus, "such that" makes sense and helps us read the sentence properly. Consider this mathematical statement:

$$\exists y \in \mathbb{R}\text{.}\ \forall x \in \mathbb{R}\text{.}\ P(x, y)$$

What would happen if we read it out loud but misplaced the phrase "such that" and used it after the "$\forall$" instead of the "$\exists$"? That would yield this sentence:

$$\exists y \in \mathbb{R} \quad \forall x \in \mathbb{R} \text{ such that } P(x, y) \text{ is True.}$$

We claim that this can be interpreted in two ways, *neither* of which is really the correct intended meaning, which is why we've written in red!

On the one hand, one might argue that such a sentence is *not* grammatical at all and has no meaning, because "such that" does not belong after a *universal* quantification. This amounts to just throwing up one's hands and saying, "I have no idea what you meant there!"

On the other hand, one might read into the sentence a little bit and argue that what the writer really meant was

$$\exists y \in \mathbb{R}, \forall x \in \mathbb{R}, \text{ such that } P(x, y) \text{ is True.}$$

or, writing out the words,

$$\text{There exists an } x \in \mathbb{R}, \text{ for each } y \in \mathbb{R}, \text{ such that } P(x, y) \text{ is } \mathsf{True}.$$

Here, the commas indicate an *inversion* of phrase order, as is common in English language. (For instance, consider the following sentence: "I laugh, at every episode of *30 Rock*, wholeheartedly." This is the same as saying "I laugh wholeheartedly at every episode of *30 Rock*.") This sentence would be equivalent, then, to writing

$$\forall x \in \mathbb{R}\text{. } \exists y \in \mathbb{R} \text{ such that } P(x, y) \text{ is } \mathsf{True}.$$

This is *not* the same as the original mathematical statement we considered and, in fact, it is actually the *other* statement we saw in the previous section (see Section 4.2.4), which was False! Recall that the other statement was similar but the phrases were reversed:

> There is a real number x such that, for every real number y, we have $y = x^3$.

which we can symbolize as

$$\exists x \in \mathbb{R}\text{. } \forall y \in \mathbb{R} \text{ such that } P(x, y) \text{ is true}$$

Look at that! The misplacement of the phrase "such that" led to a reasonable linguistic interpretation of the sentence that has the exact opposite meaning as what was originally intended. Yikes! This is why we must be careful to use "such that" *always and only* after an **existential quantification**. Remember that we will not always write that helper phrase, so you must remember to use it properly when reading a sentence to yourself in your head, or out loud to others, to make sure you have the correct, intended interpretation.

The point of this example in the previous section was to point out how important *word order* is. Now that we have symbols to replace some words and phrases, we want to emphasize how important the order of those symbols is, as well. The two mathematical statements we see above contain the exact same words and symbols, but in a different order, and one is False whereas the other is True. Clearly, order is extremely important!

4.3.3 "Fixed" Variables and Dependence

While we are on the topic of order of quantifiers, we will also mention the following example to emphasize that the order of quantifiers dictates when to consider variables as **fixed** in an expression.

Consider the statement "Any even natural number greater than or equal to 4 can be written as the sum of two primes." (Recall that this is the famous **Goldbach Conjecture** we discussed in the previous section.) To express this

statement logically and symbolically, we would write

> Let X be the set of even natural numbers, except 2.
> Let P be the set of prime numbers.
> Define $Q(n, a, b)$ to be "$n = a + b$".
> Then the claim is:
> $\forall n \in X \,.\, \exists a, b \in P \,.\, Q(n, a, b)$

Notice that we used some shorthand here. A phrase like "$\exists a, b \in P$" really means "there exists some $a \in P$ and there exists some $b \in P$", and it would be perfectly acceptable to express the above statement instead as

$$\forall n \in X \,.\, \exists a \in P \,.\, \exists b \in P \,.\, Q(n, a, b)$$

When two variables are quantified as elements from the same set, though, and the quantifications follow one another immediately in a sentence, it is very common to combine them into one quantification. We might even see mathematical statements like,

$$\forall x, y \in \mathbb{Z} \,.\, \exists a, b, c, d \in \mathbb{Z} \,.\, a + b + c + d = x + y \text{ and } a + b \neq x \text{ and } c + d \neq y$$

(What is this statement asserting, by the way? Is it True or False? Does it depend on the context of $\mathbb{Z}$? What if we used $\mathbb{N}$ or $\mathbb{R}$ instead in both places?)

Quantification "Fixes" a Variable

Look back at the example above, where we defined $Q(n, a, b)$. The reason we brough up that example was to mention that the initial quantification "$\forall n \in X$" serves to *fix* a particular value of n that will be used for the rest of the statement. After that, the assertion that "$\exists a, b \in P$" with the subsequent property $Q(n, a, b)$ depends on that *fixed, but arbitrary,* value of n.

The statement, as a whole, is saying that no matter what n is chosen, we can find values a, b that satisfy the property Q. (Notice that those values of a, b might *depend* on n, of course.) However, the *order* of quantification is telling us that those values a, b might *depend* on the chosen n. This is what we want to emphasize.

As an example, consider a particular value of the variable n in the statement. We know $8 \in X$ because 8 is even and $8 \geq 4$. What happens when $n = 8$? Can you find $a, b \in P$ such that $a + b = 8$? Sure, we can use $a = 3$ and $b = 5$. Okay, what about when $n = 14$? Can you find $a, b \in P$ that satisfy $a + b = 14$? Surely, your choices now have to be *different* than before. This is what we mean when we say a and b *depend* on n. (By the way, *can* you find a and b in this case, with $n = 14$? We can think of a couple of choices that would work!)

To make sure you're understanding this discussion, think about the following question and answer it: What is the difference between the statement above and the following one?

$$\exists n \in X \,.\, \exists a, b \in P \,.\, Q(n, a, b)$$

Is this statement True or False? Why?

4.3.4 Specifying a quantification set

Another aspect of quantifiers we want to emphasize is that we must specify a *set* whenever we use quantifiers. The sentence

$$\forall x \text{.}\ x^2 \geq 0$$

may "look true" but it is, in fact, **meaningless**. What is x? Where does it come from? "For every x ..." from where? What if x is not a number?

We *need* to specify where the object x "comes from" so that we know whether $x^2 \geq 0$ is even a well-defined, grammatical phrase, let alone whether it is True. If we amend the sentence to say

$$\forall x \in \mathbb{R} \text{.}\ x^2 \geq 0$$

then this is a well-defined, grammatical (and True!) mathematical statement. However, if we amend the sentence to say

$$\forall x \in \mathbb{C} \text{.}\ x^2 \geq 0$$

then this is a well-defined yet False mathematical statement! This is because $i \in \mathbb{C}$ but $i^2 = -1 < 0$. (Remember, we will not make significant use of the set of complex numbers $\mathbb{C}$ in this book, but it makes for interesting and enlightening examples, like this.)

The main lesson here is that **context** really matters. It can change the meaning of a statement, as well as its truth value. For this reason, we must always be sure to specify a set from which we draw variable values.

One Exception

We sheepishly admit that there is one exception to this "always specify a quantification set" rule, but there's a good reason for the exception. Consider the following claim:

> For any sets A, B, C, the equality $(A \cup B) \cap C = (A \cap C) \cup (A \cap B)$ holds true.

This is a True mathematical statement. (Can you prove it? Try using a double-containment argument!)

How might we try to write this statement in symbolic form? This is a *universal* quantification ("for any ...") so we need to use a "$\forall$" symbol. The variables here (denoted by A and B and C) are *sets*. Where do they come from? What is the set of objects we would draw them from?

We're pretty sure you're tempted to say "the set of all sets". Right? That's a big problem, though! Remember our discussion in the previous chapter about Russell's Paradox? (See Section 3.3.5 to remind yourself.) That object—the collection of all possible sets—is *not*, itself, a set! Thus, we cannot write this statement symbolically as

$$\forall A, B, C \in ___ \text{.}\ (A \cup B) \cap C = (A \cap C) \cup (A \cap B)$$

because we don't know how to fill in the blank with a *set*.

Because of this issue, we will continue to use phrases like "For any sets $A, B, C \ldots$", instead of a symbolic form. When taking notes on your own, or working out a problem on scratch paper, feel free to write "$\forall A, B, C$" and know that it really represents a quantification of sets. However, when writing more formally (say, on written homework), you should stick to the phrasing used above.

4.3.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between $\forall$ and $\exists$?

(2) How would you read the following statement out loud?

$$\forall x \in \mathbb{R}\text{.}\ \exists y \in \mathbb{R}\text{.}\ x = y^3$$

(3) Why is the following sentence **not** a proper mathematical statement?

$$\exists y\text{.}\ y + 3 > 10$$

What is the difference, if anything, between the following two statements?

$$\exists x \in \mathbb{N}\text{.}\ \exists y \in \mathbb{N}\text{.}\ x + y = 5$$

$$\exists x, y \in \mathbb{N}\text{.}\ x + y = 5$$

Are they True or False?

(4) What is the difference, if anything, between the following two statements?

$$\exists a, b \in \mathbb{Z}\text{.}\ a \cdot b = -3$$

$$\exists \heartsuit, \diamondsuit \in \mathbb{Z}\text{.}\ \heartsuit \cdot \diamondsuit = -3$$

Are they True or False?

(5) Why are the following sentences *not* properly quantified statements?

- $\exists x\text{.}\ x > 7$
- $\forall y \in \mathbb{Z}$
- $\forall z > 2\text{.}\ z^2 > 4$
- $\forall w \in \mathbb{Z}\text{.}\ w^2 = t\text{.}\ \exists t \in \mathbb{N}$

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Look back at Section 4.3.3 where we expressed the Goldbach Conjecture in symbolic notation. We defined X to be the set of all even natural numbers except 2.

Write a definition of X using symbols, with quantifiers and set builder notation (and perhaps set operations, depending on how you do it).

(2) Write an example of a mathematical statement that starts with a quantifier, and such that the statement is True if that quantifier is "$\exists$" but the statement is False if that quantifier is "$\forall$".

(3) Write an example of a variable proposition $P(x)$ such that

$$\forall x \in \mathbb{Z}\text{.}\ P(x)$$

is True but

$$\forall x \in \mathbb{N}\text{.}\ P(x)$$

is False.

(4) For each of the following mathematical statements, write it in symbolic form using quantifiers. (Be sure to properly define any variable propositions you might need, first!) Then, determine whether the statement is True or False.

(a) There is a real number that is strictly bigger than every integer.
(b) Each integer has the property that its square is less than or equal to its cube.
(c) Every natural number's square root is a real number.
(d) Every subset of $\mathbb{N}$ has the number 3 as an element.

(5) For each of the following quantified statements, say it out loud by reading the symbolic notation. Then, determine wheter the statement is True or False.

(a) $\forall x \in \mathbb{N}\text{.}\ \exists y \in \mathbb{Z}\text{.}\ x + y < 0$
(b) $\exists x \in \mathbb{N}\text{.}\ \forall y \in \mathbb{Z}\text{.}\ x + y < 0$
(c) $\exists A \in \mathcal{P}(\mathbb{Z})\text{.}\ \mathbb{N} \subset A \subset \mathbb{Z}$

(Recall that $\subset$ means "is a *proper* subset of".)

(d) Let P be the set of prime numbers.

$\forall x \in P\text{.}\ \exists t \in \mathbb{Z}\text{.}\ x = 2t + 1$

(e) $\forall a \in \mathbb{N}\text{.}\ \exists b \in \mathbb{Z}\text{.}\ \forall c \in \mathbb{N}\text{.}\ a + b < c$
(f) $\exists b \in \mathbb{Z}\text{.}\ \forall a, c \in \mathbb{N}\text{.}\ a + b < c$

4.4 Logical Negation of Quantified Statements

Let's return to those example statements we have used before. Define $P(x, y)$ to be "$y = x^3$" and then define Q_1 to be the statement

$$\text{“}\exists y \in \mathbb{R}\text{.}\ \forall x \in \mathbb{R}\text{.}\ P(x, y)\text{”}$$

and Q_2 to be

$$\text{“}\forall x \in \mathbb{R}\text{.}\ \exists y \in \mathbb{R}\text{.}\ P(x, y)\text{”}$$

Remember that Q_1 is False and Q_2 is True.

How is it that we *know* Q_1 is False? It says that there is some real number with a certain property. To declare the entire statement to be False, we might have to verify that the property does *not* hold for *every* real number y, but that would take a long time! The set $\mathbb{R}$ is infinitely large! A far more efficient approach is to show that the **negation** of this statement is True.

By "negation", we mean the **logical negation**, the statement that is the "opposite" of the original statement, in the logical sense. The logical negation of a mathematical statement has the exact opposite truth value as the original, so if we examine Q_1's negation and show it is True, then we have proven that Q_1 itself is False.

How do we negate this statement, though? We already had the right idea in mind when we noticed that we would somehow have to prove something about *every* real number y, since the original statement makes a claim of *existence*. In this section, we will explore how to properly negate statements like these.

We should note that there are some subtle, yet deep, mathematical concepts underlying what we have discussed thus far. Why is it that a mathematical statement is either True or False? Well, a cheeky (and completely correct, mind you) response would be, "Because you defined '**mathematical statement**' to be that way, silly!" Yes, indeed, we did, but *why* did we do so? What is it about the duality of True/False that is somehow *helpful* to mathematics, or *essential*? These are meaningful and difficult questions, and are definitely worth thinking about. Discussions of these topics will necessarily delve into the philoshophy of mathematics and human thought which are interesting and worthwhile pursuits, certainly, but beyond the scope and goals of this book/course. We will rely on our common, intuitive understandings of truth.

4.4.1 Negation of a universal quantification

In general, the negation of a universal (i.e. "$\forall$") claim is one of existence (i.e. "$\exists$"), and vice versa. Before we tackle the larger problem of negating any quantified statement, let's look at a simple case.

Assume S is a set and $R(x)$ is a mathematical statement, defined for every $x \in S$. The statement

$$\forall x \in S\text{.}\ R(x)$$

asserts the truth value of a variable proposition for *every* possible value of the variable x from the set S. It says that *no matter what* element x of the set S we

are referring to, we can *necessarily* conclude that the proposition $R(x)$ is true. Now, how could this statement be False, and how could we *prove* that?

If it's False that every element $x \in S$ satisfies a certain property, it must be that *at least one* element does *not* satisfy that property. To prove this, we would be expected to produce such a value; we would have to define (i.e. identify) an element x and explain why $R(x)$ does not hold for that particular element. (Think about how we understand this negation linguistically. We do this all the time in everyday language without even thinking about it.) The conclusion, then, is that the negation of the original statement is

$$\exists x \in S \text{ such that } R(x) \text{ is False}$$

We now introduce the notational symbol $\neg$ to mean **"logical negation"** or **"not"**. With this in hand we can rewrite the negated statement

$$\neg\big(\, \forall x \in S \,.\, R(x) \,\big)$$

as

$$\exists x \in S \,.\, \neg R(x)$$

The concluding phrase of that statement, $\neg R(x)$, could be simplified, depending on what $R(x)$ is.

For instance, if $S = \mathbb{R}$ and $R(x)$ is "$x^2 \geq 0$", then the negated statement would read

"$\exists x \in \mathbb{R}$ such that $x^2 < 0$"

since "$x^2 < 0$" is logically equivalent to "$\neg(x^2 \geq 0)$".

In general, though, we must leave it as "$\neg R(x)$" without knowing anything further about the proposition R. We will also point out that, in general, the phrases "$R(x)$ is False" and "$\neg R(x)$ is True" are logically equivalent; they both assert that the proposition $R(x)$ is not true.

This notion we are developing right now is what is meant by a *counterexample*, a term you have likely heard before. To *disprove* a universally quantified statement, we must prove an existentially quantified statement; that proof involves explicitly defining an element of a set that does *not* satisfy the specified property, whence the word **counterexample**.

4.4.2 Negation of an existential quantification

A statement like

$$\exists x \in S \,.\, R(x)$$

makes an existence claim. It says that there must be some element x that satisfies the property $R(x)$. To disprove a claim like this, we would seek to show that *any* value of x actually *fails* to satisfy the property R. Accordingly, we can say that the statement

$$\neg\big(\, \exists x \in S \,.\, R(x) \,\big)$$

is logically equivalent to the statement

$$\forall x \in S\text{.}\ \neg R(x)$$

This makes sense if we think about how to disprove such an existential claim. Pretend you are having a debate with a friend who told you that some kwyjibo has the property that it is a zooqa. How would you disprove him/her? You might say something like, "Nuh uh! Show me any kwyjibo you want to. I know it can't possibly be a zooqa because of the following reasons ..." and then you would explain why the property fails, no matter what.

Now, when you say "show me any" you are really performing a universal quantification! You are saying that *no matter* which kwyjibo you consider, something is true; that is, *for every* kwyjibo, or $\forall x \in K$ (where K is the set of all kwyjibos), something is True.

Think about this and consider why the logical negations we have discovered/defined make sense to you. Later on in the chapter, when we consider proof techniques, we will explain the strategy of considering an *arbitrary* kwyjibo and why this actually proves the logical negation we just wrote above. For now, we hope it is clear that

$$\forall x \in S\text{.}\ \neg R(x)$$

and

$$\exists x \in S\text{.}\ R(x)$$

have opposite truth values.

4.4.3 Negation of general quantified statements

The observations we have made so far motivate a general procedure for negating quantified statements. The statement A we defined above is of the form

$$\exists y \in \mathbb{R}\text{.}\ C(y)$$

where $C(y)$ is the rest of the statement (which *depends* on the value of y, of course). We think of $C(y)$ as some *property* of the quantified variable y; that property might have other quantifiers and variables inside it, but at a fundamental level, it is merely asserting some truth about y.

To negate this statement, we follow the method discussed above and write

$$\forall y \in \mathbb{R}\text{.}\ \neg C(y)$$

Now, we know that $C(y)$ is a universally quantified statement itself:

$$\forall x \in \mathbb{R}\text{.}\ y = x^3$$

We know how to negate that type of statement, too! That negation, $\neg C(y)$, is

$$\exists x \in \mathbb{R}\text{.}\ y \neq x^3$$

This step just uses the other negation procedure that we saw above. Then, putting it all together, we can say that $\neg A$ is the statement

$$\forall y \in \mathbb{R}\,.\; \exists x \in \mathbb{R}\,.\; y \neq x^3$$

This claim we can *prove* to be true, thus showing that the original statement must be False.

(We leave this proof as an exercise. *Hint:* Given any $y \in \mathbb{R}$, define a value of x that will force $y \neq x^3$ to be true. Notice that your choice of x will depend the value of y; how does it?)

Look at how this negation came about: we recognized that the original statement was a sequence of **nested quantifiers** (i.e. a sequence of several quantified variables in a row) with a variable proposition at the end, and we saw that we could treat part of the sequence of quantifiers as its own statement. We then "passed the negation" from the outside quantifier to the inside one, and pieced those negations together.

Following this same idea, we can figure out how to identify a statement with a longer sequence of quantifiers. For instance, look at a statement like

$$\forall a \in A\,.\; \exists b \in B\,.\; \exists c \in C\,.\; \forall d, e \in D\,.\; Q(a, b, c, d, e)$$

To start negating it, we would break off the first quantification, and treat the rest as its own proposition, $R(a)$, that depends only on a:

$$\forall a \in A\,.\; \Big(\underbrace{\exists b \in B\,.\; \exists c \in C\,.\; \forall d, e \in D\,.\; Q(a, b, c, d, e)}_{R(a)}\Big)$$

The negation can therefore be written as

$$\exists a \in A\,.\; \neg R(a)$$

but we would then have to figure out another way to write $\neg R(a)$. But hey, we would just do the same thing! We would just separate "$\exists b \in B$" from the rest and ... you see where this is going. Try working out the steps on your own, and make sure that you end up with the following as the logical negation of the original statement:

$$\exists a \in A\,.\; \forall b \in B\,.\; \forall c \in C\,.\; \exists d, e \in D\,.\; \neg Q(a, b, c, d, e)$$

In general, we can say this: To negate a statement composed *only* of quantifiers and variable propositions, just switch every "$\forall$" to "$\exists$", and vice-versa, and negate the propositions. Don't alter any of the sets over which we quantify, merely the quantifiers themselves and the ensuing propositions; it wouldn't make sense to change the universe of discourse. Later on, we will look at how to negate other types of statements, more complicated ones built from other connectives. Before we do that, we need to move on and define and discuss those other connectives.

4.4.4 Method Summary

Let's summarize what we have discovered in this section.

- **Negating a universal quantification:**

  Let X be a set and let $P(x)$ be a proposition. Then the negation of a universal quantification, like this,

  $$\neg\big(\forall x \in X\,.\,P(x)\big)$$

  is written as

  $$\exists x \in X\,.\,\neg P(x)$$

  In words, we have shown that saying

  > It is *not* the case that, for every $x \in X$, $P(x)$ holds.

  is equivalent to saying

  > There exists an element $x \in X$ such that $P(x)$ fails.

- **Negating an existential quantification:**

  Let X be a set and let $Q(x)$ be a proposition. Then the negation of an existential quantification, like this,

  $$\neg\big(\exists x \in X\,.\,Q(x)\big)$$

  is written as

  $$\forall x \in X\,.\,\neg Q(x)$$

  In words, we have shown that saying

  > It is *not* the case that there exists an $x \in X$ such that $Q(x)$ holds.

  is equivalent to saying

  > For every element $x \in X$, $Q(x)$ fails.

Don't Change the Quantification Set!

We mentioned above that it wouldn't make sense to change the universe of discourse when negating a statement. To think about why this makes sense, take a real-life example.

Suppose we said "Every book on this bookshelf is written in English." How would you prove to us that we are lying, that our statement is actually False? You would have to produce a book *on this shelf* that is written in a different language. You couldn't bring in a French novel from the room down the hall and say, "See, you were wrong!" That wouldn't prove anything about the claim we made; the realms of discourse are different, and we didn't make any claim

about what's going on in any bookshelves in other rooms. We only asserted something about this *particular* shelf.

For the same reason, when negate a statement like

$$\forall b \in T \,.\, P(b)$$

we obtain

$$\exists b \in T \,.\, \neg P(b)$$

without changing that realm of discourse, the set T. The original claim only asserts something about elements of T, so its negation does only that, as well.

4.4.5 The Law of the Excluded Middle

You know what? Let's actually discuss *why* we can talk about a statement and its **logical negation**. Built into our definition of **mathematical/logical statement** is the idea that such a sentence must have exactly one truth value, either True or False. Why can we do this? Well, we're in charge of the definitions here! Mathematicians have to set the ground rules—the **axioms**—of their systems and we want our logical system to ensure that every claim we make is decidedly True or False, and not both, and not neither.

This dichotomy is truly an **axiom** of our system. It is widely adopted in most of mathematics, and is famously known as **The Law of the Excluded Middle**. The name comes from this very idea, that every claim is True or False, so there is no *middle ground* between those two sides; that middle is excluded.

In essence, this makes what we do in mathematics fruitful: every claim has a truth value, and our goal is to find that truth value. Sometimes, though, we have to fall back on this axiom, this law we agreed upon, and just *guarantee* that some claim is either True or False, without knowing *which* truth value actually applies. We present an interesting and striking example of this idea here.

Proposition 4.4.1. *There exist real numbers a and b that are both irrational such that a^b is rational.*

(Remember that a **rational number** is one that can be expressed as a fraction of integers, and an **irrational number** is a real number that is *not* rational. Can you think of some examples of both rational and irrational numbers?)

Proof. We know $\sqrt{2}$ is irrational. (Question: Why? Can you prove that? Try it now. We will prove this very soon, as well!)

The number $\sqrt{2}^{\sqrt{2}}$ is either rational or irrational. (This is where the Excluded Middle is used.) Let's consider these two cases separately.

- Suppose that the number $\sqrt{2}^{\sqrt{2}}$ is, indeed, rational. Then $a = \sqrt{2}$ and $b = \sqrt{2}$ is the example we seek, because a and b are both irrational and a^b is rational.

- Now, suppose that the number $\sqrt{2}^{\sqrt{2}}$ is irrational. In this case, we can use $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$ as the example we seek, because a and b are both irrational and

$$a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = (\sqrt{2})^2 = 2$$

  which is rational.

In either case, we have found an example of real numbers $a, b \in \mathbb{R}$ such that both a and b are irrational and yet a^b is rational. This proves the claim. □

This is an example of a **non-constructive** proof. It tells us something exists (and narrows it down to two possibilities, even) without actually telling us *exactly* which possibility is the one we sought all long. It is a direct use of the Law of the Excluded Middle that causes this.

(Question: Can you prove somehow that $\sqrt{2}^{\sqrt{2}}$ is irrational? It is True, but there is no known "simple" proof of this fact. Maybe you can find one!)

Most of the proofs we do here will be of the **constructive** variety (but not all). These might be more satisfying to you, and we're inclined to agree. If we claim something exists, we should be able to *show* it to you, right? If we just talked for a while about *why* some such object exists somewhere else, without being able to point to it, you would have to believe us, but you might not feel great about it. Constructive proofs are *subjectively better* because of this, and we will always strive for one when we can. Sometimes, though, a constructive proof is not immediately clear, and we have to make a non-constructive one, like we did here.

4.4.6 Looking Back: Indexed Set Operations and Quantifiers

Look back at Section 3.6.2, where we defined set operations—union and intersection, in Definitions 3.6.3 and 3.6.4, respectively—performed over index sets. The main idea was that we could express the union/intersection of an entire class of sets all at once using a shorthand notation.

Look carefully at those definitions. What characterized whether an object actually *is* an element of an indexed union, for example? That object needed to be an element of *at least one* of the constituent sets of the union. That is, there needed to *exist* some set of which that obejct is an element. This sounds like an *existential quantification*, doesn't it?

Likewise, what characterized whether an object is an element of an indexed intersection? That object needed to be an element of *all* the constituent sets. That is, *for all* of those sets, that object must be an element thereof. This is a *universal quantification*.

With those observations now made, we can rewrite those definitions of indexed set operations using our new quantifier notation:

Definition 4.4.2. *Suppose I is an index set and $\forall i \in I \bullet A_i \subseteq U$, for some universal set U. Then*

$$\bigcup_{i \in I} A_i = \{x \in U \mid \exists k \in I \bullet x \in A_k\}$$
$$\bigcap_{i \in I} A_i = \{x \in U \mid \forall i \in I \bullet x \in A_i\}$$

Try working with some of the examples and exercises in that Section 3.6.2 again. Do the definitions make more sense now?

4.4.7 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the **negation** of a mathematical statement? How are a statement and its negation related?

(2) Why is the negation of a $\forall$ claim an $\exists$ one?

Why is the negation of an $\exists$ claim a $\forall$ one?

(3) What is a non-constructive proof? To what type of claim—$\exists$ or $\forall$—does this term apply?

(4) Consider the claim

$$\forall x \in S \bullet P(x)$$

Why is its negation *neither* of the following?

$$\forall x \notin S \bullet P(x)$$

$$\exists x \notin S \bullet \neg P(x)$$

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) For each of the following statements, write its negation. Which one—the original or the negation—is True?

(a) $\forall x \in \mathbb{R}.\ \exists n \in \mathbb{N}.\ n > x$

(b) $\exists n \in \mathbb{N}.\ \forall x \in \mathbb{R}.\ n > x$

(c) $\forall x \in \mathbb{R}.\ \exists y \in \mathbb{R}.\ y = x^3$

(d) $\exists y \in \mathbb{R}.\ \forall x \in \mathbb{R}.\ y = x^3$

(2) For each of the following statements, write its negation. Which one—the original or the negation—is True?

(a) $\exists S \in \mathcal{P}(\mathbb{N}).\ \forall x \in \mathbb{N}.\ x \in S$

(b) $\forall S \in \mathcal{P}(\mathbb{N}).\ \exists x \in \mathbb{N}.\ x \in S$

(c) $\forall x \in \mathbb{N}.\ \exists S \in \mathcal{P}(\mathbb{N}).\ x \in S$

(d) $\exists x \in \mathbb{N}.\ \forall S \in \mathcal{P}(\mathbb{N}).\ x \in S$

(3) Let $I = \{x \in \mathbb{R} \mid 0 < x < 1\}$.

For each of the following defined sets, write out the defining condition that determines whether a number $y \in \mathbb{R}$ is an element of the set, using quantifiers.

Then, determine what the set is, and write your answer using set-builder notation.

(Try to *prove* your claim, as well, using a double-containment argument!)

(a)
$$S = \bigcup_{x \in I} \{y \in \mathbb{R} \mid x < y < 2\}$$

(b)
$$T = \bigcap_{x \in I} \{y \in \mathbb{R} \mid -x < y < x\}$$

(c)
$$V = \bigcup_{x \in I} \{y \in \mathbb{R} \mid -3x < y < 4x\}$$

(4) Let $P = \{y \in \mathbb{R} \mid y > 0\}$. Consider this statement:

$$\forall \varepsilon \in P.\ \exists \delta \in P.\ \forall x \in \{y \in \mathbb{R} \mid -\delta < y < \delta\}.\ |x^3| < \varepsilon$$

Write out the logical negation of this statement.

What does this statement say? What does its negation say?

Which one is True? Can you prove it?

(5) Let A, B, C, D be arbitrary sets.

Let $P(x), Q(x, y), R(x, y, z)$ be arbitrary variable propositions.

Write the negation of each of the following statements.

(a) $\forall a \in A\,.\; \exists b \in B\,.\; Q(a,b)$
(b) $\forall a \in A\,.\; \neg P(a)$
(c) $\forall c \in C\,.\; \forall d \in D\,.\; \neg Q(c,d)$
(d) $\exists a_1, a_2 \in A\,.\; \forall d \in D\,.\; R(a_1, a_2, d)$
(e) $\forall b_1, b_2, b_3 \in B\,.\; \neg R(b_1, b_2, b_3)$
(f) $\exists b \in B\,.\; \forall c \in C\,.\; \forall d \in D\,.\; R(d, b, c)$

4.5 Logical Connectives

To build mathematical statements from simpler ones (meaning ones composed of just quantifiers and propositions) we can connect several statements with certain words and phrases—such as "and", "or", and "implies"—to create more complicated statements and assert further claims and truths. We call these words and phrases **logical connectives**, and each of them has their own corresponding mathematical symbol and meaning. These meanings will make sense to you, based on our intuitive grasp of the English language and rational thought, but we emphasize that one of the major goals of introducing mathematical logic and its corresponding notation is to build these intuitions into rigorous and unambiguous concepts.

Throughout this section, let us assume that P and Q are arbitrary mathematical statements. These statements themselves can be composed of complicated combinations of quantifiers and other connectives and all sorts of mathematical notions. The point is that the way we combine P and Q into a larger statement is independent of their individual compositions. Before, we saw that "$\neg(\forall x \in X\,.\; R(x))$" is equivalent to "$\exists x \in X\,.\; \neg R(x)$", regardless of what the statement $R(x)$ was and how complicated it might have been. This idea continues here. We can talk about how to combine two statements without knowing what they are, individually.

We should also point out that these constituent statements, P and Q, may actually be variable propositions. For instance, we will consider how to connect two variable propositions, $P(x)$ and $Q(x)$, that each depend on some variable x. The definitions and methods we develop in this section apply to these variable propositions even though these propositions, themselves, do not have truth values without being told what the value of the variable x is.

When we want to talk about those propositions meaningfully and mathematically, we will have to **quantify** the variable x. Thus, if we have variable propositions $P(x)$ and $Q(x)$, we can still meaningfully define $P(x) \wedge Q(x)$ (where $\wedge$ means "and" as you'll see in the next section). We could then, in an example or a problem, talk about a claim of the form

$$\exists x \in X\,.\; P(x) \wedge Q(x)$$

This is a mathematical **statement**.

Essentially, the point we want to make is that these connectives still apply to variable propositions, but the relevant variables will have to be quantified

somewhere in an overall statement to make the variable proposition into a proper **mathematical statement**.

4.5.1 And

To say

"P and Q" is True

means that *both* statements have the truth value: True. If either one of the statements P or Q were False, then the statement "P and Q" would be False, as well. This definition encapsulates this idea:

Definition 4.5.1. *We use the symbol "$\wedge$" between two mathematical statements to mean "and". For instance, we read "$P \wedge Q$" as "P and Q".*

This is referred to as the ***conjunction*** *of P and Q.*

The truth value of "$P \wedge Q$" is True *when both P and Q are true, and the truth value is* False *otherwise.*

Here are some examples to demonstrate this definition:

Example 4.5.2.

$$\begin{array}{ll} (1+3=4) \wedge (\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0) & \textsf{True} \\ (1+3=5) \wedge (\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0) & \textsf{False} \\ (1+3=5) \wedge (\exists x \in \mathbb{Q}\text{.}\ x^2 = 2) & \textsf{False} \end{array}$$

Notation: Parentheses

It's sometimes common to drop the parentheses that we used in the examples above. For example, the first line in the above example can be written equivalently as

$$1+3=4 \ \wedge\ \forall x \in \mathbb{R}\text{.}\ x^2 \geq 0$$

Using the parentheses tends to make the statement more readable. Without them, we have to think for a few extra moments about where one part of the statement ends and the next one begins, but we can still eventually make sense of it. We will try to use parentheses whenever they make a statement more easily understandable.

Notation: Sets and Logic

You might notice the similarity between the logical connective "$\wedge$" and the set operator "$\cap$". This is not a coincidence!

As we will discuss below in Section 4.5.4, we can write the definition of "$\cap$" using the connective "$\wedge$" because of the underlying logic of that set operator. Try it now, and then glance ahead to that section briefly, if you'd like. In general, though, be careful to keep these two notations separate! If A and B are sets, the phrase "$A \wedge B$" is not well-defined; what was meant is "$A \cap B$".

4.5.2 Or

To say

"P or Q" is True

means that "P is True, or Q is True". We need to know that *one* of the statements is True to declare that the entire statement has the truth value True. We don't care whether *both* P and Q are true or not, merely that *at least one* of them is true.

This is in contrast with the so-called "exclusive OR" of computer science, also known as XOR, which declares "P XOR Q" to be False when both P and Q are True. In mathematics, we use the **inclusive "or"**. We only care whether at least one of the statements holds.

Definition 4.5.3. *We use the symbol $\vee$ between two mathematical statements to mean "or". For instance, we read "$P \vee Q$" as "P or Q".*

This is referred to as the **disjunction** *of P and Q.*

The truth value of "$P \vee Q$" is True *when* **at least one** *of P and Q is* True *(even when both are* True*), and the truth value is* False *otherwise.*

Example 4.5.4.

$$(1+3=4) \vee (\forall x \in \mathbb{R}.\ x^2 \geq 0) \qquad \text{True}$$
$$(1+3=5) \vee (\forall x \in \mathbb{R}.\ x^2 \geq 0) \qquad \text{True}$$
$$(1+3=5) \vee (\exists x \in \mathbb{R}.\ x^2 < 0) \qquad \text{False}$$

Notation

The same notes about notation that we mentioned in the previous subsection apply here, as well. First, the use of parentheses—like in the above examples—is helpful but not technically required. We will try to use them whenever it helps, though.

Second, you might notice the similarity between the logical connective "$\vee$" and the set operator "$\cup$". Again, this is not a coincidence! Try rewriting the definition of "$\cup$" using "$\vee$", and glance ahead briefly at Section 4.5.4. In general, though, be careful to keep these two notations separate! If A and B are sets, the phrase "$A \vee B$" is not well-defined; what was meant is "$A \cup B$".

4.5.3 Conditional Statements

This is the hardest logical connective to work with and continually gives students some problems, so we want to be extra careful and clear about this one. We want the statement **"If P, then Q"** (sometimes written as "P implies Q") to have the truth value True when the truth of Q *necessarily* follows from the truth of P. That is, we want this statement to be True if the following holds:

> Whenever P is True, Q is *also* True.

Truth Table and Definition

Since this is the hardest connective to suss out, semantically, let's introduce the idea of a **truth table** to make the notation easier:

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \implies Q$	$Q \implies P$
T	T	F	T	T	T	T
T	F	F	F	T	F	T
F	T	T	F	T	T	F
F	F	T	F	F	T	T

You may have seen a truth table before, in other mathematics courses, but even if you haven't, don't worry! Here's the main idea: Each column corresponds to a particular mathematical statement and its corresponding truth values. Each row corresponds to a particular *assigment* of truth values to the constituent statements, P and Q.

Notice that there are 4 rows because P and Q can each have one of two different truth values, so there are 4 possible combinations of those choices. Reading across a particular row, we find the corresponding truth values for other statements, based on what the T and F assignments for P and Q are in the first two columns.

Notice that the columns for $P \wedge Q$ and $P \vee Q$ follow the definitions given above. The column for $P \wedge Q$ has only one T, and it corresponds to the case where *both* P and Q are True. All other cases make $P \wedge Q$ a False statement. Likewise, the column for $P \vee Q$ has only one F, and it corresponds to the case where *both* P and Q are False. All other cases make $P \vee Q$ a True statement.

Now, why are the last two columns the way they are? Let's say that I make the claim "If you work hard, then you will get an A in this course". Here, P is "You work hard" and Q is "You will get an A". When can you call me a *liar*? When can you declare I told the truth? Certainly, if you work hard and get an A, I told the truth. Also, if you work hard and don't get an A, then I lied to you. However, if you don't work hard, then no matter what happens, *you don't get to call me a liar!* My claim didn't cover your situation; I was assuming all along you would just work hard! Thus, I didn't speak an untruth and so, by the Law of the Excluded Middle, I *did* speak the truth. The negation of a lie is a truth.

This situation—where the third and fourth rows of the $P \implies Q$ column are True—is known as a **false hypothesis**. When the statement on the left of the " $\implies$ " does not hold, then we are not in a situation that is meant to be addressed by the claim, so we cannot assert that the claim is False. Therefore, the claim must be True (again, by the Law of the Excluded Middle).

Let's make the proper definition of this symbol and then consider more examples to illustrate the definition.

Definition 4.5.5. *We use the symbol " $\implies$ " between to mathematical statements to mean "If . . . then" or "implies". For instance, we read "$P \implies Q$" as "If P, then Q" or "P implies Q".*

This is referred to as a **conditional statement**.

The truth value of "$P \implies Q$" is True *assuming that Q holds whenver P holds.*

The truth value is False only *in the case where P is* True *and yet Q is* False.

We refer to P as the **hypothesis** *of the conditional statement and Q as the* **conclusion**.

That key word "whenever" in the definition should indicate to you why the *false hypothesis* cases make sense. When we know P is true and can deduce that Q is also true, then we get to declare $P \implies Q$ as True. If P wasn't true to begin with, we cannot declare $P \implies Q$ to be False. We only get to say $P \implies Q$ is false when Q did not necessarily follow from P, i.e. when there is an instance where the hypothesis P is True but the conclusion Q is False.

Examples

Here are several examples to help you get the idea:

$$\begin{aligned}
(1+3=4) &\implies (\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0) && \text{True}\\
(1+3=5) &\implies (\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0) && \text{True}\\
(1+3=5) &\implies (\text{Abraham Lincoln is alive}) && \text{True}\\
(1+1=2) &\implies (0=1) && \text{False}\\
(0=0) &\implies (\exists x \in \mathbb{R}\text{.}\ x^2 < 0) && \text{False}\\
(\text{Pythagorean Theorem}) &\implies (1=1) && \text{True}\\
(0=1) &\implies (1=1) && \text{True}
\end{aligned}$$

Notice that the second and third examples are True because they have the false hypothesis "$1 + 3 = 5$". Regardless of the conclusion, the entire conditional statement must be True because of this. It doesn't matter that "$\forall x \in \mathbb{R}\text{.}\ x^2 \geq 0$" happens to be True or that "Abraham Lincoln is alive" happens to be False; that false hypothesis determines the statement's truth value.

Also, notice that the second-to-last example is True, but it doesn't help us determine whether or not the Pythagorean Theorem itself is True! This is what we did in the False "spoof" of the theorem back in Chapter 1. Look back to Section 1.1.1, specficially "Proof 2". We assumed the Pythagorean Theorem was True and then logically derived a True statement from that assumption. That does not make the hypothesis valid, just because we derived a valid conclusion!

This idea is so important, that we will even show you again this striking example. Notice that it's logical form is exactly the same as that other spoof:

"Proof". Assume $1 = 0$. Then, by the symmetric property of $=$, it is also true that $0 = 1$. Adding these two equations tells us $1 = 1$, which is True. Therefore, $0 = 1$. □

This is the main point here:

> Knowing a conditional statement, overall, is True, doesn't tell us *anything* about the truth values of the constituent propositions.

This is also strikingly illustrated in the third and seventh statements above; both conditional claims are True, but we certainly don't get to conclude that Abraham Lincoln is alive, or that $0 = 1$.

"Implies" is not the same as "Can be deduced from"

There is often some confusion with using the word "implies" to mean an "If ... then ..." statement, a conditional statement. We belive this arises because of some connotations surrounding the word "implies"; specifically, it seems to convey some sort of *causality*. For instance, consider this statement:

$$1 + 3 = 4 \implies 2 + 3 = 5$$

This is a True conditional statement, and our minds probably recognize this because we can just take the hypothesis, namely $1 + 3 = 4$, and add 1 to both sides, yielding the equation in the conclusion. In this sense, it seems that the truth of the hypothesis has some influence on the truth of the conclusion: we can deduce one *directly* from the other. This does not have to be the case, in general!

Look back at the first example given above:

$$(1 + 3 = 4) \implies (\forall x \in \mathbb{R}.\ x^2 \geq 0)$$

What does the fact that $1+3 = 4$ have to do with the fact that any real number squared is non-negative? Does it even have any connection? We don't actually care! We can still identify this conditional statement as True, whether or not we can find a way to deduce the conclusion directly *from* the hypothesis (and whether or not such a deduction even exists). Only the truth values of the constituent statements matter.

Granted, when we work on proving conditional statements, we will likely try to deduce one statement directly from another. It's important to keep in mind, though, that this is a consequence of our proof strategy, and not an underlying part of how conditional statements are defined. For these reasons, we tend to write conditional statements using the "If ... then ..." form, instead of using "implies". We might sometimes use it, and we're sure you'll see it in other mathematical writings. For now, though, we'll try to avoid it as much as possible, while we're still learning about these logical statements and connectives.

Quantifying Variables: Still Important!

In mathematics, we often want to prove conditional statements that involve variables. For instance, we might want to show that, in the context of the real numbers $\mathbb{R}$, the following conditional claim holds:

$$x > 1 \implies x^2 - 1 > 0$$

That sentence, written in the line above, is itself a **variable proposition**, and the definition of the symbol " $\implies$ " applies to it.

If we knew that $x > 1$ and also $x^2 - 1 > 0$, then we could declare the conditional to be True. If we knew that $x \leq 1$, then we wouldn't even care whether $x^2 - 1 > 0$ or not; we could declare the conditional to be True. This is how the definition of " $\implies$ " applies here.

Remember, though, that the conditional claim, as written above, is not technically a mathematical statement. We were making that claim in the context of the real numbers, so it would really make sense to write

$$\forall x \in \mathbb{R}\text{.}\ \big(x > 1 \implies x^2 - 1 > 0\big)$$

This is, ultimately, what the writer was trying to claim, so they should just say so! We make the same recommendation to you. These logical connectives—$\wedge$ and $\vee$ and $\implies$ —make sense and can be applied to variable propositions. Outside of that scope, somewhere else in the statement you're putting together, there must be some kind of quantification on those variables. Only then can we be assured that the sentence is a mathematical statement with one truth value.

Writing " $\implies$ " using "$\vee$"

There is a useful and important idea worth mentioning. This is partly because we will use it later, but also partly because it might help you understand conditional statements and learn how to use them.

This idea hinges on the notion of a *false hypothesis*. Consider a conditional statement, $P \implies Q$. If P fails, then the entire statement is True, regardless of Q's truth value. However, if P holds, then we definitely need Q to hold, as well, to declare the entire statement to be True.

These observations allow us to identify two ways in which a conditional statement can hold, and write these two ways in an "or" statement. Either $\neg P$ holds (i.e. a false hypothesis), or else Q holds. In either of these situations, the conditional statement $P \implies Q$ must hold! Let's state this claim outright for you to consider:

> The conditional statement "$P \implies Q$" and the statement "$\neg P \vee Q$" have the same truth value.

This is a good example of a **logical equivalence**, which is a topic we will discuss in the next section. For now, we will present the truth table for the two statements mentioned above. Notice that they have the same truth value, regardless of the truth values of the constituent statements, P and Q. This serves as further verification that the statements are equivalent, in addition to the description we provided above.

P	Q	$\neg P$	$\neg P \vee Q$	$P \implies Q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Investigating More Examples

Let's consider some more examples of conditional claims and decide whether they are True or False. In so doing, we are helping you to better understand how $\implies$ works.

Then, we'll move on to proof *strategies* and discuss how to formally and rigorously prove claims like these, with logical connectives and quantifiers.

Example 4.5.6. We will start here with a "real world" example, to get used to the logic involved. Throughout this example, pretend we are in a class that only has lectures officially scheduled on Mondays, Wednesdays, and Fridays.

You'll notice that we will take two statements, P and Q, and consider all four possible combinations of these statements and their negations to make conditional statements.

- "If we have lecture today, then it is a weekday."

  (Note: There is some *implicit quantification* in this sentence. We are really saying that "For all days d in the weekly calendar, if we have lecture on day d, then d is a weekday." We think that the main idea is conveyed more succinctly by the sentence above, so we will use that version. Keep in mind that this is the meaning of the sentence and, in our discussion below, we will have to consider the different cases of that quantification.)

  This can be written logically by defining P to be "We have lecture today" and Q to be "Today is a weekday"; then the claim is $P \implies Q$.

  Is this True? Notice that the statements P and Q don't specify *what day it is*, so if we were to assert this claim to be True, that truth should be *independent* of the current day. That is, *whatever* day it is, we would have to show that $P \implies Q$ holds. Let's do that by considering a few cases:

  - Pretend today is Saturday or Sunday. Since we never have lecture on these days, the conditional claim is not a lie, so it is True.
  - Pretend today is Monday or Wednesday or Friday. If we do, indeed, have lecture today, then it is definitely a weekday, so the claim is True.
  - Pretend today is Tuesday or Thursday. We don't typically have lecture today, but even in the special case that we did (for some rescheduling reason), it would still be a weekday, so the claim is True.

  In any of the possible cases, the claim holds. Thus, $P \implies Q$ is True.

  You might interject to say, "Why bother with the cases at all? Couldn't we just say that no matter what day it is, supposing we have lecture, then we conclude it is definitely a weekday?" Well, yes, we could! That's actually a better strategy, a more *direct* route, you might say.

  This hints at how we will prove conditional claims in the future. Since we don't, in fact, care about the situations where we don't have lecture (the

false hypothesis) we only need to *suppose* we have lecture on day X and *deduce* that X is a weekday. This is the method we will use to **directly prove** a conditional claim.

- "If it is a weekday, then we have lecture today."

  This is logically written as $Q \implies P$, using the same definitions of P and Q as the previous example.

  Is this claim True? Definitely not! We didn't have lecture on the first Tuesday of the semester, yet that day was a weekday. Thus, the claimer lied in that instance! On that Tuesday, Q was True but P was False. Thus, $Q \implies P$ is False.

- "If it is not a weekday, then we don't have lecture today."

  This is logically written as $\neg Q \implies \neg P$, using the same definitions of P and Q as the example above.

  Is this claim True? Yes, and we can prove it directly. Suppose that it is not a weekday; that is, pretend today is Saturday or Sunday. Obviously, the university would not be so sadistic as to schedule a lecture on a weekend, so we can necessarily declare that we don't have lecture, i.e. $\neg P$ holds. This shows that $\neg Q \implies \neg P$ is a True statement.

  (Question: Why did we not need to consider the case where today is a weekday?)

- "If we don't have lecture today, it is not a weekday."

  This is logically written as $\neg P \implies \neg Q$, using the same definitions of P and Q as the example above.

  Is this claim True? Well, let's think about it. What if we pretend we don't have lecture today. What can we necessarily say? Is it definitely not a weekday? I don't think so! Maybe it's a Tuesday, and we just don't have a scheduled lecture. This shows that the claim is False; we have an instance where the hypothesis ($\neg P$) holds, and yet the conclusion ($\neg Q$) does not hold.

  Notice that there *are* situations where $\neg P$ holds *and* $\neg Q$ does, as well. For instance, if today were Saturday, then certainly we don't have lecture *and* it is not a weekday. This *specific instance* does not meant that the claim is True, though! We need to verify its truth for *all instances*.

Example 4.5.7. Let's do the same kind of analysis with a more "mathy" example. Throughout this example, let A and B be arbitrary sets. Also, let P be "$A \subseteq B$", and let Q be "$A - B = \varnothing$".

We will do what we did in the previous example and consider all four possible ways of combining P and Q and their negations to make conditional statements.

- Is $P \implies Q$ True?

Yes! Let's pretend A and B satisfy the relationship $A \subseteq B$. This means *every* element of A is also an element of B. Therefore, we cannot possibly have an element of A that does *not* belong to B, Since $A - B$ is the set of elements that belong to A and not B, we conclude that there are no such elements, so $A - B = \varnothing$.

- Is $Q \implies P$ True?

  Yes! Let's pretend $A - B = \varnothing$. This means there are *no* elements of A that are also not elements of B. (Think about this.) Put another way, any element $x \in A$ *cannot* have the property that $x \notin B$ (or else $x \in A - B$ and so $A - B \neq \varnothing$); thus, $x \in B$, necessarily. Hey, this is exactly what $A \subseteq B$ means! Whenever $x \in A$, we conclude $x \in B$, too. This shows $Q \implies P$ is true.

- Is $\neg Q \implies \neg P$ True?

  Hmm, this is harder to figure out. Let's pretend $\neg Q$ holds; this means $A - B \neq \varnothing$. That is, there exists some element x that satisfies $x \in A$ and $x \notin B$. Certainly, then $A \not\subseteq B$, because we have identified an element of A that is not an element of B (whereas the $\subseteq$ relationship would dictate that every element of A is also an element of B). Thus, $\neg Q \implies \neg P$ is True.

- Is $\neg P \implies \neg Q$ True?

  Again, we need to think about this. Let's just write down what $\neg P$ means. To say $A \not\subseteq B$ means there exists some element $x \in A$ that also satisfies $x \notin B$. (This is what we used in the previous case, too.) Okay, what does that tell us? Consider the set $A - B$. Does it have any elements? Yes, it has at least x as an element! Since $x \in A \wedge x \notin B$, we can say $x \in A - B$. Thus, $A - B \neq \varnothing$, so we conclude that $\neg P \implies \neg Q$ is True.

Observations and Facts About " $\implies$ "

Okay, now we have some practice working with conditional statements and determining their truth values. What you should notice from the examples we discussed is that knowing $P \implies Q$ holds does **not** tell us anything about $Q \implies P$. In both of these examples above, $P \implies Q$ was True; however, $Q \implies P$ was True in one example and False in the other. There is nothing we can necessarily say, with certainty, about $Q \implies P$, even if we know the truth value of $P \implies Q$. This idea is so important, that we will touch on it in the next subsection.

For now, let's make a few more remarks about the " $\implies$ " connective.

- Remember that, given mathematical statements P and Q, the sentence "$P \implies Q$" is, itself, another mathematical statement. It has a truth value. That truth value *depends* on P and Q (in the way we defined it

above), but it does not tell us *anything* about the truth values of P and Q. So, if you just write down the claim

$$\text{Blah blah} \quad \Longrightarrow \quad \text{Yada yada}$$

on your paper, we have no idea if you're trying to assert whether "Blah blah" or "Yada yada" are True or False! To a mathematician, this just says:

> The conditional statement " 'Blah blah' implies 'Yada yada' " is True.

If you wish to make some kind of inference or deduction, then use some helping words and sentences to indicate that. Write something like this:

> $P \implies Q$ because ...
>
> Also, P holds because ...
>
> Therefore, Q holds.

If you have studied formal logic before, or have seen this type of argument in a philosophy course, then you might recognize this as **Modus Ponens**.

- The idea of a **false hypothesis** yielding a True conditional is kind of weird; we realize this. It's a direct consequence of the Law of the Excluded Middle. Under a false hypothesis, we don't get to say the overall statement is False, so it must be True, since it must be one or the other.
- Remember that we can always write a conditional statement without the " $\Longrightarrow$ " symbol by converting it to an "or" statement.

  > The statements "$P \implies Q$" and "$\neg P \vee Q$" always have they always have the same truth value.

Converse and Contrapositive

Let's give some names to the different types of conditional statements related to a given conditional statement. We will refer to these frequently later on.

Definition 4.5.8. *Let P and Q be mathematical statements. Consider the "original" claim, $P \implies Q$.*

We refer to $Q \implies P$ as the **converse** *of the original claim.*

We refer to $\neg Q \implies \neg P$ as the **contrapositive** *of the original claim.*

By our observations in the previous subsection, we know that the **converse** does not *necessarily* have the same truth value as the original. What we will see (and prove) in the next section is that the **contrapositive** always has the *same* truth value as the original claim. (This is the notion of **logical equivalence**, which we will discuss in full detail in the next section.)

You might wonder why we need this terminology at all. Well, since the contrapositive can be shown to be *logically equivalent* to the original claim, it gives rise to a valid proof method when we try to prove conditional statements. We will develop that very soon. That's why we use the contrapositive.

The converse is interesting because its truth value is not necessarily tied to that of the original statement: even knowing the original is True, the converse might be True and it might be False. Thus, whenever we prove a claim $P \implies Q$ to be true, a mathematician (probably) immediately wonders, "Hmm, does the converse also hold?" It's a natural question to ask, and worth thinking about whenever you face a conditional statement. (In fact, if you ever find yourself at a party with mathematicians, and you hear someone talking about an "If ... then ..." statement, you should ask, "Does the converse hold, as well?" You might impress your fellow guests.)

The converse is also the subject of a common logical fallacy you might encounter in everyday life. Perhaps you are trying to argue that $A \implies B$ in a debate with a friend. What if they retort, "Well, B doesn't necessarily imply A! Your argument is wrong!" Have you ever been frustrated by that situation? You might have been tempted to shout, "So what? I wasn't trying to say anything about whether $B \implies A$. I was talking about $A \implies B$, you ..." (We'll cut ourselves off before we get mean.) Whether or not your friend is right, knowing the truth value of the converse doesn't tell you anything about the truth value of your original claim. You should tell them that! Next time, just say, "You're talking about the converse, which is not necessarily logically connected to my claim."

Okay, now that we have defined all of the requisite logical symbols and seen some examples, it's time to move on and start proving things about them! But first, a short aside about set operations, and then a few practice problems.

4.5.4 Looking Back: Set Operations and Logical Connectives

Look back at Sections 3.4 and 3.5, where we defined subsets and set operations. All of those definitions made use of some logical ideas, but we wrote them in English back then, relying on our collective intuition and working knowledge of logic. We can rewrite them now using quantifiers and connectives!

First, recall the definition of **subset**. We write $A \subseteq B$ if the following holds: whenever $x \in A$, we can also say $x \in B$. Notice that key word, "whenever", which signals both a *universal quantification* and a *conditional statement*. Think about how you would rewrite the definition of $A \subseteq B$, using this notion, then read on to see our version ...

Definition 4.5.9. *Let A, B, U be sets, where $A, B \subseteq U$ (i.e. U is a universal set). We say A is a* **subset** *of B, and write $A \subseteq B$, if and only if*

$$\forall x \in U \bullet x \in A \implies x \in B$$

This makes sense because it asserts that "whenever" statement we wrote in the above paragraph: whenever $x \in A$, we must also be able to conclude $x \in B$; "if $x \in A$, then $x \in B$" must hold.

Look back again at the definitions of the set operations we gave. Try to write your own versions of those definitions using logical symbols and then read ours here. Think about how they make sense, how they express the same underlying ideas.

Definition 4.5.10. *Let A, B, U be sets, where $A, B \subseteq U$ (i.e. U is a universal set). Then,*

$$\begin{aligned} A \cap B &= \{x \in U \mid x \in A \ \wedge \ x \in B\} \\ A \cup B &= \{x \in U \mid x \in A \ \vee \ x \in B\} \\ A - B &= \{x \in U \mid x \in A \ \wedge \ \neg(x \in B)\} = \{x \in U \mid x \in A \ \wedge \ x \notin B\} \\ \overline{A} &= \{x \in U \mid \neg(x \in A)\} = \{x \in U \mid x \notin A\} \end{aligned}$$

While we're here, we can also redefine a **partition** of a set. This will make use of logical connectives, but this also hearkens back to indexed sets and how they are defined in terms of quantifiers. Everything we've learned is coming together here!

Definition 4.5.11. *Let A be a set. A* **partition** *of A is a collection of sets that are pairwise disjoint and whose union is A.*

That is, a partition is formed by an index set I and non-empty sets S_i (defined for every $i \in I$) that satisfy the following conditions:

(1) $\forall i \in I \bullet S_i \subseteq A$.

(2) $\forall i, j \in I \bullet i \neq j \implies S_i \cap S_j = \varnothing$.

(3) $\bigcup_{i \in I} S_i = A$

Look back to Definition 3.6.9 to see how we originally defined a partition. Do you see how we are saying the same thing here, just using logical symbols?

4.5.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between $\wedge$ and $\vee$?

(2) What is the difference between $\wedge$ and $\cap$?

What is the difference between $\vee$ and $\cup$?

(3) Write out a truth table for the statement $P \implies Q$.

(4) Why are $P \implies Q$ and $\neg P \vee Q$ logically equivalent statements?

(5) What is the converse of a conditional statement?

What is the contrapositive of a conditional statement?

(6) Are the truth values of a conditional statement and its converse related?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) For each of the following sentences, rewrite it using logical notation and determine whether it is True or False.

(a) Every integer is either strictly positive or strictly negative.

(b) For any given real number, there exists a natural number that is strictly bigger.

(c) For every real number, if it is negative, then its cube is also negative.

(d) There is a subset of $\mathbb{Z}$ that has the property that whenever a number is an element of that set, so is its square.

(e) There is a natural number that is both even and odd.

(2) Rewrite each of the following $\forall$ claims as a conditional statement, and determine whether it is True or False.

(a) $\forall x \in \{y \in \mathbb{N} \mid \exists k \in \mathbb{Z}\text{.}\ y = 2k\}\text{.}\ x^2 \in \{y \in \mathbb{N} \mid \exists k \in \mathbb{Z}\text{.}\ y = 2k\}$

(b) $\forall x \in \{y \in \mathbb{N} \mid \exists k \in \mathbb{Z}\text{.}\ y = 2k+1\}\text{.}\ x^2 \in \{y \in \mathbb{N} \mid \exists k \in \mathbb{Z}\text{.}\ y = 2k+1\}$

(c) $\forall t \in \{z \in \mathbb{R} \mid z^2 > 4\}\text{.}\ t > 2$

(3) Rewrite each of the following conditional statements as a $\forall$ claim, using set-builder notation, and determine whether it is True or False.

(a) $\forall x \in \mathbb{R}\text{.}\ x > 3 \implies x^2 < 9$

(b) $\forall x \in \mathbb{R}\text{.}\ x < 3 \implies x^2 < 9$

(c) $\forall t \in \mathbb{R}\text{.}\ t^2 - 6t + 9 \geq 0 \implies t \geq 3$

(4) Let's define the following variable propositions:

$$P(x) \text{ is } \text{“ } \frac{1}{2} < x \text{ ”}$$
$$Q(x) \text{ is } \text{“ } x < \frac{3}{2} \text{ ”}$$
$$R(x) \text{ is } \text{“ } x^2 = 4 \text{ ”}$$
$$S(x) \text{ is } \text{“ } x + 1 \in \mathbb{N} \text{ ”}$$

For each of the following statements, determine whether it is True or False.

(a) $\forall x \in \mathbb{N}\text{.}\ P(x)$

(b) $\forall x \in \mathbb{N}\text{.}\ Q(x) \implies P(x)$

(c) $\forall x \in \mathbb{Z}\text{.}\ Q(x) \implies P(x)$

(d) $\exists x \in \mathbb{N}\text{.}\ \neg S(x) \vee R(x)$

(e) $\exists x \in \mathbb{Z}\text{.}\ R(x) \wedge \neg S(x)$

(f) $\forall x \in \mathbb{R}\text{.}\ R(x) \implies S(x)$

(g) $\exists x \in \mathbb{R}\text{.}\ P(x) \wedge S(x)$

(h) $\forall x \in \mathbb{Z}\text{.}\ R(x) \implies \big(P(x) \vee Q(x)\big)$

(5) For each of the following conditional statements, write it in logical notation, and use this write its converse and its contrapositive; then, determine the truth values of all three: the original statement, the converse, and the contrapositive.

(a) If a real number is strictly between 0 and 1, then so is its square.

(b) If a natural number is even, then so is its cube.

(c) Whenever an integer is a multiple of 10, it is a multiple of 5.

4.6 Logical Equivalence

In this section, the major goal is to introduce the idea of **logical equivalence** and prove a few fundamental claims. Essentially, we want to decide when some complicated logical statements are actually "the same", in the sense of truth values. Since mathematical statements may depend on some propositional variables, we might not be able to conclude anything specific about their truth values. However, we can sometimes prove that two mathematical statements will have *the same truth value*, for all possible values of the variables they contain. That's a really nice conclusion to make! We get to say that they have the same truth value, regardless of what it is. In that sense, we are proving the two statements to be **equivalent**, in a logical sense.

4.6.1 Definition and Uses

The following definition introduces a convenient symbol for the notion of logical equivalence described in the above paragraph:

Definition 4.6.1. *Let P and Q be mathematical statements. We use the symbol "$\iff$" to mean "is* **logically equivalent** *to", or "has the same truth value as".*

That is, we write "$P \iff Q$" when P and Q always have the same truth value, regardless of whether it is T *or* F*.*

We read "$P \iff Q$" aloud as "P is logically equivalent to Q" or "P **if and only if** *Q".*

This type of statement is known as a **biconditional** *(or a* **bi-implication***).*

Let's take the truth tables we saw in the last section and add a new column for this symbol:

P	Q	$\neg P$	$\neg P \vee Q$	$P \implies Q$	$Q \implies P$	$P \iff Q$
T	T	F	T	T	T	T
T	F	F	F	F	T	F
F	T	T	T	T	F	F
F	F	T	T	T	T	T

In the column for $P \iff Q$, an entry has the truth value T when (and only when) P and Q have the same truth value. This happens in Row 1, where both are T, and in Row 4, where both are F. Notice, then, that $P \iff Q$ has the truth value T if and only if

$$(P \implies Q) \wedge (Q \implies P)$$

is a True statement. This is the notion of **logical equivalence**: $P \iff Q$ means that both $P \implies Q$ and $Q \implies P$ hold. Whatever truth value P has, Q is guaranteed to have the same truth value, and vice-versa:

- Supposing that P is True, then $P \implies Q$ tells us that Q must also be True.
- Supposing that P is False, then $Q \implies P$ tells us that Q *cannot* be True (since $Q \implies P$ would be False, in that case), so Q must also be False.

Either way, P and Q have the same truth value.

Examples

Example 4.6.2. Look again at the third and fourth columns in the truth table above. They prove the following logical equivalence:

$$(P \implies Q) \iff (\neg P \vee Q)$$

Whatever the truth value of $P \implies Q$ (which, of course, depends on P and Q), it must be the same as the truth value of $\neg P \vee Q$. We've mentioned this equivalence before, and we will make use of it fairly often in the future.

Example 4.6.3. Look at this truth table:

P	Q	$\neg P$	$\neg Q$	$P \implies Q$	$\neg Q \implies \neg P$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Regardless of the truth values of P and Q, we find that $P \implies Q$ and $\neg Q \implies \neg P$ have the *same* truth values. Thus, they are *logically equivalent*, and we can write:

$$(P \implies Q) \iff (\neg Q \implies \neg P)$$

This is the fact that we stated (without proof) in the previous section:

> The contrapositive of a conditional statement is logically equivalent to the original statement.

A different proof of this fact makes use of the way to express a conditional statement as a disjunction. Recall the logical equivalence

$$\big(P \implies Q\big) \iff \big(\neg P \vee Q\big)$$

that we mentioned in the previous example. Now, think about the contrapositive of that original conditional statement:

$$\neg Q \implies \neg P$$

Applying the same disjunctive form to that statement yields the following equivalence:

$$\big(\neg Q \implies \neg P\big) \iff \big(\neg(\neg Q) \vee \neg P\big)$$

Now, noticing that $\neg(\neg Q)$ is equivalent to just Q, and remembering that the order of a disjunction is irrelevant (i.e. $P \vee Q$ and $Q \vee P$ have the same truth value) we find that

$$\big(\neg Q \implies \neg P\big) \iff \big(\neg P \vee Q\big) \iff \big(P \implies Q\big)$$

This shows, in another way, that a conditional statement and its contrapositive have the same truth value!

Example 4.6.4. Later in this section, we will prove the following logical equivalences, which hold no matter what the propositions P and Q and R are:

$$\begin{aligned}
\neg(P \wedge Q) &\iff \neg P \vee \neg Q \\
(P \wedge Q) \wedge R &\iff P \wedge (Q \wedge R) \\
P \vee (Q \wedge R) &\iff (P \vee Q) \wedge (P \vee R) \\
\neg(P \implies Q) &\iff P \wedge \neg Q
\end{aligned}$$

Each of these is an assertion that the expressions on both sides of the $\iff$ symbol have the same truth value. Can you see why these claims are True? Can you think of how to prove them?

If and Only If

Logical equivalence has a nice relationship with the phrase "if and only if". To say "P if and only if Q" means we are asserting that both "P if Q" and "P only if Q" hold. One of these corresponds to $P \implies Q$ and the other corresponds to $Q \implies P$, so asserting both are true means exactly what we have described:

$$P \iff Q \quad \text{is the same as saying} \quad (P \implies Q) \wedge (Q \implies P)$$

Now, which one is which, though? When we say "P if Q", this means "If Q, then P." That is,

$$\text{“}P \text{ if } Q\text{”} \quad \text{is the same as saying} \quad Q \implies P$$

Sussing out the other direction is a little harder! What does "P *only if* Q" really mean? This sentence is asserting that, in a situation where P holds, it must also be the case that Q holds. That is, knowing P holds means we also immediately know Q holds. Put even another way, whenever P is true, we necessarily know that Q is true. This is the same as saying $P \implies Q$ holds!

Another way of thinking about it is as follows. Saying "P only if Q" is the same as saying it cannot be the case that P holds and Q does not. Written logically, we have

$$\neg\big(P \wedge \neg Q\big)$$

Later on in this section, we will state and prove **DeMorgan's Laws for Logic**. One of those laws tells us how to negate that statement inside the parentheses. (You might already know these logical laws, in fact. If not, you can glance ahead at Sections 4.6.5 and 4.6.6 for a preview.) The conclusion is:

$$\neg P \vee Q$$

Hey look, that's logically equivalent to $P \implies Q$, as we observed already! Cool. Just further confirmation that "P only if Q" means $P \implies Q$.

Using "$\iff$" in Definitions

We will also often use the "$\iff$" symbol in a **definition** to indicate that the term defined is an equivalent term for the property that is used in the definition. For example:

$$\text{We say } x \in \mathbb{Z} \text{ is } \textbf{even} \iff \exists k \in \mathbb{Z}\textbf{.}\ x = 2k$$

That is, the notion of an integer being even is equivalent to knowing that the number is twice an integer. Similarly, we can define **odd**:

$$\text{We say } x \in \mathbb{Z} \text{ is } \textbf{odd} \iff \exists k \in \mathbb{Z}\textbf{.}\ x = 2k + 1$$

These are formal definitions, mind you, and are the only way of guaranteeing an integer is even (or odd). We will use these definitions soon to rigorously prove some facts about integers and arithmetic. Every time we want to assert a particular integer (call it x) is even, we need to show there exists an integer k that satisfies $x = 2k$. That is, we have to *satisfy the definition* by appealing to the logical equivalence given in the definition.

Biconditional Statements: A Technical Distinction

We can also use the symbol "$\iff$" to express two conditional statements at once. Technically speaking, this is not *exactly* the same as asserting a logical equivalence, but it conveys a similar idea, so we allow the symbol to be used in both ways.

A logical equivalence involves some undefined propositions, and it asserts that the two statements will have the same truth value, regardless of the truth values of those propositions. For instance,

$$(P \implies Q) \iff (\neg P \vee Q)$$

is a perfect example of a logical equivalence. Without telling you what P and Q are, we can't be sure exactly what $P \implies Q$ and $\neg P \vee Q$ mean. However, we don't need to know what P and Q are to know that those two statements definitely have the same truth value.

The situation is slightly different when the two statements on either side of the "$\iff$" are actually proper mathematical statements, with no undefined propositions. For instance, consider this statement:

$$\forall x \in \mathbb{R}\text{.}\ \ (x > 0) \iff \left(\frac{1}{x} > 0\right)$$

This is a logical claim, and it asserts that, whenever x is a real number, knowing one of those two facts—$x > 0$ or $\frac{1}{x} > 0$—necessarily guarantees the other. That is, if I told you I have a real number in mind and it is positive, you get to conclude that its reciprocal is positive, too. Also, if I told you I have a real number in mind whose reciprocal is positive, you would get to conclude that the number itself is positive, too. It goes *both ways*. (Question: What if I told you I had a *negative* real number in mind? Could you conclude anything about its reciprocal? Why or why not?)

Do you see how this is different? Given an arbitrary $x \in \mathbb{R}$, the statement "$x > 0$" is decidedly either True or False. Its truth value isn't left undetermined. This distinguishes it from the example given above, where the truth value of the individual statements is unknown, yet we can still declare those truth values must be identical.

For lack of a better, widespread term for these kinds of statements, we will refer to them as **biconditionals**. This is because they are really meant to represent two conditional statements that go "in opposite directions":

$$\forall x \in \mathbb{R}\text{.}\ \left[\left((x > 0) \implies \left(\frac{1}{x} > 0\right)\right) \wedge \left(\left(\frac{1}{x} > 0\right) \implies (x > 0)\right)\right]$$

This is what the statement above says: each part of the statement implies the other one.

This term is not necessarily standard in other mathematical writing, but we wanted to point out this technical distinction so you are aware of it. You might approach a mathematicial logician or set theorist and use the phrase "logical

equivalence", and they might be confused or take offense to the way in which you use it. This is not a big worry, mind you! Since we are learning about these fundamental ideas now for the first time, we don't necessarily have to keep in mind all of the technical details that lie underneath these concepts. Also, in the remainder of this book, we might use "logical equivalence" and "biconditional" interchangeably. This is fine and acceptable for now.

The main point behind using the "$\Longleftrightarrow$" symbol is to assert that two statements *have the same truth value*. The only difference between a "logical equivalence" and a "biconditional" is whether or not those statements contained therein have any arbitrary, undefined propositions. This is a minor distinction to be made, in the grand scheme of things, so we will consider it only briefly here.

4.6.2 Necessary and Sufficient Conditions

There are two terms occasionally used in mathematics that convey the two directions of a biconditional statement: **necessary** and **sufficient**. They correspond exactly to the "only if" and "if" parts of the biconditional. These terms are motivated by the natural types of questions mathematicians ask.

Sufficient: P, if Q

If we identify some interesting fact or property—call it P—of a mathematical object, we might wonder, "When can we *guarantee* such a property holds? Is there some condition we can check that will give us a 'Yes' answer right away?" This is what a **sufficient** condition is, a property that guarantees P will also hold. It is "sufficient" in the sense that it is "enough" to conclude P; we don't need any other outside information.

Let's say we have identified a proposition Q as a sufficent condition for P. How can we express this logically? Well, knowing Q is sufficient to conclude P, so we can easily write this as a conditional statement:

$$Q \implies P \qquad \text{means } Q \text{ is a } \textbf{sufficient} \text{ condition for } P$$

Said another way, this conditional statement expresses: "P, if Q".

Necessary: P only if Q

We also might wonder, "How can we guarantee that P fails? Is there some condition we can check that will tell us this right away?" This is what a **necessary** condition is, a property that is necessary or *essential* for the property P to hold. This condition is not necessarily enough to conclude that P holds, but for P to even have a chance of holding, this condition better hold, too.

Think about the logical connections here. Say we have established a property Q that is a **necessary** condition for P. How can we express the relationship between P and Q, symbolically? That's right, we can use a conditional statement. Knowing P holds tells us that Q definitely holds; it was necessary for P

to be true. This is expressed as

$$P \implies Q \qquad \text{means } Q \text{ is a } \textbf{necessary} \text{ condition for } P$$

Said another way, this conditional statement expresses: "P only if Q".

We could also think of this in terms of the contrapositive. If Q does not hold, then P cannot hold, either. That is,

$$\neg Q \implies \neg P$$

which is the contrapositive of the conditional statement above, $P \implies Q$. We know these are logically equivalent forms of the same statement.

Examples

Example 4.6.5. Let $P(x)$ the proposition "x is an integer that is divisible by 6". For each of the following conditions, let's identify whether it is a **necessary** or **sufficient** condition (or possibly both!) for $P(x)$ to hold.

(1) Let $Q(x)$ be "x is an integer that is divisible by 3".

To determine whether $Q(x)$ is a necessary condition, let's assume $P(x)$ holds. Can we deduce $Q(x)$ holds, too? Well, yes! To say an integer x is divisible by 6 means that it is divisible by both 2 and 3. Thus, it is certainly divisible by 3, so $Q(x)$ holds.

To determine whether $Q(x)$ is a sufficient condition, let's assume $Q(x)$ holds. Can we deduce $P(x)$ holds, too? Hmm ... knowing that x is an integer divisible by 3, is it also *definitely* divisible by 2, allowing us to conclude it is divisible by 6? We think not! Consider $x = 3$; notice $Q(3)$ holds but $P(3)$ does not.

This shows $Q(x)$ is only a necessary condition, and not a sufficient one.

(2) Let $R(x)$ be "x is an integer that is divisible by 12."

Following similar reasoning to the above example, we can conclude that $R(x)$ is a sufficient condition for $P(x)$, but not a necessary one (because we can have $x = 6$, where $P(6)$ holds but $R(6)$ does not hold).

(3) Let $S(x)$ be "x is an integer such that x^2 is divisible by 6".

We'll let you work with this one on your own ... Is $S(x)$ a necessary condition for $P(x)$? Is it a sufficient one?

Be careful, and notice that we specified x, itself, is an integer ...

4.6.3 Proving Logical Equivalences: Associative Laws

Now, let's actually **prove** some logical equivalences! In doing so, we will be working on our ability to read and understand and write logical statements using quantifiers and connectives. We will also be developing some fundamental

logical results that we can apply in the near future to develop proof techniques. These techniques will be the foundation of the rest of our work, and everything else we do will involve implementing some combination of these proof strategies and logical concepts.

Let's start with some of the simpler symbolic logical laws. Showing something is *associative* essentially means we can "move around the parentheses" willy-nilly and end up with the same thing. You probably use the fact that addition is associative all the time! To add x to $y+z$, we can just add z to $x+y$ instead and know we get the same answer. That is, we can rest assured that

$$x + (y + z) = (x + y) + z$$

We can *move* the parentheses around wherever we want to and so, ultimately, we can just pretend as if they aren't even there and just write

$$x + y + z$$

because the order in which we interpret the two additions is irrelevant. The same kind of result applies to conjuctions and disjunctions of logical statements, and that's what we will prove now.

Theorem 4.6.6. *Let P, Q, R be logical statements. Then*

$$P \wedge (Q \wedge R) \iff (P \wedge Q) \wedge R$$

and

$$P \vee (Q \vee R) \iff (P \vee Q) \vee R$$

We will actually prove these claims in two separate ways: (1) via truth tables, and (2) via semantics (i.e. words). They are both perfectly valid, but we want to show you both of them to let you decide which style you like better.

Proof 1. First, we will prove these claims via truth tables. Observe the table for conjunctions:

P	Q	R	$P \wedge Q$	$Q \wedge R$	$P \wedge (Q \wedge R)$	$(P \wedge Q) \wedge R$
T	T	T	T	T	T	T
T	T	F	T	F	F	F
T	F	T	F	F	F	F
T	F	F	F	F	F	F
F	T	T	F	T	F	F
F	T	F	F	F	F	F
F	F	T	F	F	F	F
F	F	F	F	F	F	F

Thus, $P \wedge (Q \wedge R) \iff (P \wedge Q) \wedge R$ because their corresponding columns are identical, in every case.

Next, observe the table for disjunctions:

P	Q	R	$P \vee Q$	$Q \vee R$	$P \vee (Q \vee R)$	$(P \vee Q) \vee R$
T	T	T	T	T	T	T
T	T	F	T	T	T	T
T	F	T	T	T	T	T
T	F	F	T	F	T	T
F	T	T	T	T	T	T
F	T	F	T	T	T	T
F	F	T	F	T	T	T
F	F	F	F	F	F	F

Thus, $P \vee (Q \vee R) \iff (P \vee Q) \vee R$ because their corresponding columns are identical, in every case. □

Proof 2. Second, let's prove these claims by analyzing them, semantically. Consider the first claim,

$$P \wedge (Q \wedge R) \iff (P \wedge Q) \wedge R$$

To show that the two sides are *logically equivalent*, we need to show both of the following conditional statements are True:

$$P \wedge (Q \wedge R) \implies (P \wedge Q) \wedge R$$

and

$$(P \wedge Q) \wedge R \implies P \wedge (Q \wedge R)$$

($\implies$) Let's prove the first conditional statement. Suppose $P \wedge (Q \wedge R)$ is True. This means P is True and $Q \wedge R$ is True. By definition, this means P is True and Q is True and R is True. Certainly, then $P \wedge Q$ is True and R is True, by definition. Thus, $(P \wedge Q) \wedge R$ is True, as well.

($\impliedby$) Now, let's prove the second conditional statement. Suppose $(P \wedge Q) \wedge R$ is True. This means $P \wedge Q$ is True and R is True. By definition, this means P is True and Q is True and R is True. Certainly, then P is True and $Q \wedge R$ is True, by definition. Thus, $P \wedge (Q \wedge R)$ is True, as well.

Since we have shown both conditional statements, we conclude the two sides are, indeed, logically equivalent.

Next, consider the second claim of the theorem,

$$P \vee (Q \vee R) \iff (P \vee Q) \vee R$$

To show that the two sides are *logically equivalent*, we need to show both of the following conditional statements are True:

$$P \vee (Q \vee R) \implies (P \vee Q) \vee R$$

and

$$(P \vee Q) \vee R \implies P \vee (Q \vee R)$$

($\implies$) Let's prove the first conditional statement. Suppose $P \vee (Q \vee R)$ is True. This means either P is True or $Q \vee R$ is True. This gives us two cases.

1. Suppose P is True. This means $P \vee Q$ is True, by definition. Thus, $(P \vee Q) \vee R$ is True, also by definition.

2. Suppose $Q \vee R$ is True. This means either Q is True or R is True. Again, this gives us two cases.

   (a) Suppose Q is True. This means $P \vee Q$ is True, by definition. Thus, $(P \vee Q) \vee R$ is True, also by definition.

   (b) Suppose R is True. This means $(P \vee Q) \vee R$ is True, by definition.

In any case, we find that $(P \vee Q) \vee R$ is True. Thus, this conditional statement is True.

($\Longleftarrow$) Let's prove the second conditional statement. Suppose $(P \vee Q) \vee R$ is True. This means either $P \vee Q$ is True or R is True. This gives us two cases.

1. Suppose $P \vee Q$ is True. This means either P is True or Q is True. This gives us two cases.

   (a) Suppose P is True. This means $P \vee (Q \vee R)$ is true, by definition.

   (b) Suppose Q is True. This means $Q \vee R$ is True, by definition. Thus, $P \vee (Q \vee R)$ is true, also by definition.

2. Suppose R is True. This means $Q \vee R$ is True, by definition. Thus, $P \vee (Q \vee R)$ is True, also by definition.

In any case, we conclude that $P \vee (Q \vee R)$ is True. Thus, this conditional statement is True.

Since we have shown both conditional statements hold, we conclude the two sides are, indeed, logically equivalent. □

Okay, what have we accomplished with these proofs? What have we proven, and how? Why did it work?

Let's mention a consequence of these proofs, before going on to discuss and compare the proffs, themselves. We proved that the "$\wedge$" and "$\vee$" connectives are associative, so the order in which we evaluate parenthetical statements involving only one such connective does not matter. For example, we now know that "$P \wedge (Q \wedge R)$" has the same meaning as "$(P \wedge Q) \wedge R$". Accordingly, in the future, we will just write these statements without the parentheses: "$P \wedge Q \wedge R$".

Reflecting: Truth Tables vs. Semantics

Let's talk about the truth tables first. Since P, Q, and R are logical statements, they are each, individually, True or False. The eight rows of the truth tables consider all possible assignments of truth values to those three constituent statements. The first three columns tell us whether P, Q, R are True or False. The next two columns correspond to the more complicated constituent parts of the logical statements in the claim, and the last two columns correspond to the two parts of the actual claim in the theorem. By comparing those last two columns,

we can decide whether or not those two statements are logically equivalent. (Remember that "logically equivalent" means "has the same truth value as, no matter the assignment of truth values to P and Q and R". Thus, observing that the two columns have identicals entries, row by row, is sufficient to show that the two statements are logically equivalent.)

Next, let's talk about the semantic proofs. How do you feel about them? They were certainly longer, right? Disregarding that, though, did they feel like good proofs? Were they clear? Correct, even? Reread the proofs above and think about these questions. We will emphasize that they are fully correct proofs. The use of cases is essential when trying to prove a disjunction (an "*or*" statement) holds. When we suppose something is True and deduce that something else is True, that's how we prove a conditional statement is True. We will further analyze these techniques very soon, but we hope that giving you a first example like this will help you later on.

For the remainder of this section, we will use a truth table to verify simple claims like these. The proofs are much shorter that way! We are sure that you can go through the details of a semantic proof, like the ones we gave above, if you need further convincing or extra practice with interpreting symbolic logical claims as English sentences.

4.6.4 Proving Logical Equivalences: Distributive Laws

In arithmetic, you've used the fact that multiplication **distributes** across addition. That is, we know that

$$\forall x, y, z \in \mathbb{R}\text{.}\ x \cdot (y + z) = x \cdot y + x \cdot z$$

Hey, look, we wrote this in symbolic notation! Do you see why it says what you already know about the distributive property?

Here we will state and prove two similar laws. They will tell us that the logical connectives "$\wedge$" and "$\vee$" also distribute across each other.

Theorem 4.6.7. *Let P, Q, and R be logical statements. Then*

$$P \wedge (Q \vee R) \iff (P \wedge Q) \vee (P \wedge R)$$

and

$$P \vee (Q \wedge R) \iff (P \vee Q) \wedge (P \vee R)$$

Proof. We will use truth tables to verify these two claims. Observe, for the first

claim, that

P	Q	R	$Q \vee R$	$P \wedge Q$	$P \wedge R$	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	T	T	T	T	T
T	T	F	T	T	F	T	T
T	F	T	T	F	T	T	T
T	F	F	F	F	F	F	F
F	T	T	T	F	F	F	F
F	T	F	T	F	F	F	F
F	F	T	T	F	F	F	F
F	F	F	F	F	F	F	F

Notice that the last two columns are identical, thus proving the desired logical equivalence.

Observe, for the second claim, that

P	Q	R	$Q \wedge R$	$P \vee Q$	$P \vee R$	$P \vee (Q \wedge R)$	$(P \vee Q) \wedge (P \vee R)$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	T	F	F	F
F	F	T	F	F	T	F	F
F	F	F	F	F	F	F	F

Again, notice that the last two columns are identical, thus proving the desired logical equivalence. □

4.6.5 Proving Logical Equivalences: De Morgan's Laws (Logic)

Let's prove some logical equivalences involving **negations**. The following two laws are named for the British mathematician **Augustus De Morgan**. He is credited with establishing these logical laws, as well as introducing the term **mathematical induction**! We are grateful and indebted to his work in mathematics.

De Morgan's Laws for Logic state some logical equivalences about negating conjunctions and disjunctions.

Theorem 4.6.8. *Let P and Q be logical statements. Then*

$$\neg\big(P \wedge Q\big) \iff \neg P \vee \neg Q$$

and

$$\neg\big(P \vee Q\big) \iff \neg P \wedge \neg Q$$

Proof. We prove the first claim by a truth table:

P	$\neg P$	Q	$\neg Q$	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P \vee \neg Q$
T	F	T	F	T	F	F
T	F	F	T	F	T	T
F	T	T	F	F	T	T
F	T	F	T	F	T	T

And then we prove the second claim by a truth table:

P	$\neg P$	Q	$\neg Q$	$P \vee Q$	$\neg(P \vee Q)$	$\neg P \wedge \neg Q$
T	F	T	F	T	F	F
T	F	F	T	T	F	F
F	T	T	F	T	F	F
F	T	F	T	F	T	T

□

These two laws are extremely useful! In fact, we can use them to prove similar statements about sets.

4.6.6 Using Logical Equivalences: DeMorgan's Laws (Sets)

The following statements "look a lot like" the statements in DeMorgan's Laws for Logic we saw above. We will see exactly why they look so similar when we see the proof!

Theorem 4.6.9. *Let A and B be any sets, and suppose that $A, B \subseteq U$, so the complement operation is defined in the context of U. Then,*

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

and

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

We will prove these claims using logical equivalences and DeMorgan's Laws for Logic. Our method will be to show that, in either case, the property of being an element of the set on the left-hand side of the equation is logically equivalent to being an element of the set on the right-hand side. This proves both parts of a double-containment argument in one fell swoop.

Proof. Let's prove the first set equality. Let $x \in U$ be arbitrary and fixed. Then,

$$\begin{aligned}
x \in \overline{A \cup B} &\iff x \notin A \cup B && \text{Definition of complement} \\
&\iff \neg\,(x \in A \cup B) && \text{Definition of } \notin \\
&\iff \neg\,[(x \in A) \vee (x \in B)] && \text{Definition of } \cup \text{ and } \vee \\
&\iff \neg(x \in A) \wedge \neg(x \in B) && \text{DeMorgan's Law for Logic} \\
&\iff (x \notin A) \wedge (x \notin B) && \text{Definition of } \notin \\
&\iff (x \in \overline{A}) \wedge (x \in \overline{B}) && \text{Definition of complement} \\
&\iff x \in \overline{A} \cap \overline{B} && \text{Definition of } \wedge \text{ and } \cap
\end{aligned}$$

Remember that " $\wedge$" is a *logical* operation, while "$\cap$" is a *set* operation. We had to be careful about which one made sense in every sentence we wrote. Also, notice that we used DeMorgan's Law for Logic in the middle of the proof, to convert a negation of a disjunction into the conjunction of two negations.

This chain of logical equivalences shows that

$$x \in \overline{A \cup B} \iff x \in \overline{A} \cap \overline{B}$$

so, in the context of U, the property of being an element of $\overline{A \cup B}$ is logically equivalent to the property of being an element of $\overline{A} \cap \overline{B}$. Therefore,

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Let's prove the second equality now, with a similar method. Let $x \in U$ be arbitrary and fixed. Then,

$$\begin{aligned}
x \in \overline{A \cap B} &\iff x \notin A \cap B && \text{Definition of complement} \\
&\iff \neg\,(x \in A \cap B) && \text{Definition of } \notin \\
&\iff \neg\,[(x \in A) \wedge (x \in B)] && \text{Definition of } \cap \text{ and } \wedge \\
&\iff \neg(x \in A) \vee \neg(x \in B) && \text{DeMorgan's Law for Logic} \\
&\iff (x \notin A) \vee (x \notin B) && \text{Definition of } \notin \\
&\iff (x \in \overline{A}) \vee (x \in \overline{B}) && \text{Definition of complement} \\
&\iff x \in \overline{A} \cup \overline{B} && \text{Definition of } \vee \text{ and } \cup
\end{aligned}$$

This chain of logical equivalences shows that

$$x \in \overline{A \cap B} \iff x \in \overline{A} \cup \overline{B}$$

so, in the context of U, the property of being an element of $\overline{A \cap B}$ is logically equivalent to the property of being an element of $\overline{A} \cup \overline{B}$. Therefore,

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Voilà! We have proven both equalities stated in the theorem. □

Notice the striking similarities between the two proofs. They used exactly the same method, and the only real difference is flipping a "$\cap$" to a "$\cup$", and vice-versa. Because we proved something about how to do this already (DeMorgan's Laws for Logic), we can cite that result and make this proof short and sweet. Wouldn't you agree this is far easier and more concise than writing out a full double-containment proof for these two claims? (Try it!)

4.6.7 Proving Set Containments via Conditional Statements

Whenever you can, go ahead and use the method we used in the previous section, with DeMorgan's Laws for Logic and Sets; that is, feel free to prove set

relationships via conditional statements and logical equivalences. In general, when you're proving an equality, you need to be sure that all of your claims really are "$\iff$" claims. In the previous section, we only applied definitions and a theorem about logical equivalences, so we were positive about all of the directions of the "$\iff$" arrows in the proof. Whenever you write a proof like this, read over it again once you're done and ask yourself at every line, "Does this actually work? Does the implication work both ways here?"

Let's see another example of this technique in action. It will be slightly more complicated, in that we have to define some variable propositions because the claim given is not fundamentally identical to DeMorgan's Laws for Logic. We will, though, invoke a *logical* law that we just proved, and use it to establish a *sets* law.

Proposition 4.6.10. *Let A, B, C be any sets, with $A, B, C \subseteq U$, where U is a universal set. Then,*

$$A \cap (B - C) = (A \cap B) - C$$

Much like the previous example, DeMorgan's Laws for Sets, we will establish a logical equivalence between being an element of the left-hand side and being an element of the right-hand side. (Again, this is like proving both sides of a double-containment proof all at once.) To do this, we will just establish some variable propositions that refer to the properties of being an element of A, B, and C, respectively. From there, the result will follow from a logical law.

Proof. Let A, B, C be any sets, with $A, B, C \subseteq U$, where U is a universal set. We define the following variable propositions:

$$\begin{aligned} &\text{Let } P(x) \text{ be "} x \in A\text{"} \\ &\text{Let } Q(x) \text{ be "} x \in B\text{"} \\ &\text{Let } R(x) \text{ be "} x \in C\text{"} \end{aligned}$$

Let $x \in U$ be arbitrary and fixed. With these definitions, we can write the following chain of logical equivalences (where "Defn" is just space-saving shorthand for "Definition"):

$$\begin{aligned} x \in A \cap (B - C) &\iff x \in A \wedge (x \in B - C) && \text{Defn of } \cap \\ &\iff x \in A \wedge (x \in B \wedge x \notin C) && \text{Defn of } - \\ &\iff P(x) \wedge (Q(x) \wedge \neg R(x)) && \text{Defn of } P(x), Q(x), R(x), \notin \\ &\iff (P(x) \wedge Q(x)) \wedge \neg R(x) && \text{Associative Law for } \wedge \\ &\iff (x \in A \wedge x \in B) \wedge x \notin C && \text{Defn of } P(x), Q(x), R(x) \\ &\iff x \in A \cap B \ \wedge \ x \notin C && \text{Defn of } \cap \\ &\iff x \in (A \cap B) - C && \text{Defn of } - \end{aligned}$$

This shows that

$$x \in A \cap (B - C) \iff x \in (A \cap B) - C$$

holds True for any element x in the universe U. Therefore,

$$A \cap (B - C) = (A \cap B) - C$$

□

Think about why we needed to make sure all of these claims are truly *if and only if* statements. We are ensuring that any element x that is an element of a set on one side of the equality is also necessarily an element of the set on the other side; but, furthermore, we are ensuring that any element x that is *not* an element of one set is *also* not an element of the other set. The biconditional statements "go both ways", so we prove both the "is an element of" and "is *not* an element of" parts of the claim all at once.

To illustrate our previous warnings, consider the following claim as an example of a proof where one "direction" of a $\iff$ claim *fails*.

Proposition 4.6.11. *Let X, Y, Z be any sets, with $X, Y, Z \subseteq U$, for some universal set U. Then, the following containment holds:*

$$(X \cup Y) - Z \subseteq X \cup (Y - Z)$$

You might recognize this claim as Problem 3.11.17! In that problem, we asked you to prove this claim using a containment argument, taking an arbitrary $x \in U$ and supposing it is an element of the left-hand side set, then deducing it must also be an element of the right-hand side set. We will do (essentially) the same thing here, but the argument will be recast in logical terms and symbols. We will do this to (1) give us more practice with making these types of arguments, but also (2) to recognize precisely *where* in the argument the "reverse" direction *fails*. Remember that, in Problem 3.11.17, we also asked you to find an example that shows that the $\supseteq$ direction is not *necessarily* True. This means that the logical argument working in that direction would break down somewhere. We will see precisely where that is, and we can use it to help us come up with that required counterexample.

Proof. Let X, Y, Z be any sets, with $X, Y, Z \subseteq U$, for some universal set U. Let $x \in U$ be arbitrary and fixed. We can write the following chain of logical equivalences:

$$\begin{aligned}
x \in (X \cup Y) - Z &\iff x \in X \cup Y \ \wedge \ x \notin Z && \text{Defn of } - \\
&\iff (x \in X \ \vee \ x \in Y) \wedge x \notin Z && \text{Defn of } \cup \\
&\iff (x \in X \ \wedge \ x \notin Z) \ \vee \ (x \in Y \ \wedge \ x \notin Z) && \text{Distr. Law}
\end{aligned}$$

Scratch work:

From here, what further logical equivalences could we assert? We could simplify the right-hand side and express it as

$$x \in X - Z \vee x \in X - Z$$

and, therefore,

$$x \in (X - Z) \cup (Y - Z)$$

This is not what the claim was, but this procedure so far would be a valid proof of a *different* claim, namely that

$$(X \cup Y) - Z = (X - Z) \cup (Y - Z)$$

However, our right-hand side is

$$X \cup (Y - Z)$$

but we are not trying to prove an equality, merely a *containment*. Thus, the goal of the rest of our proof is to prove this conditional claim:

$$\Big((x \in X \wedge x \notin Z) \vee (x \in Y \wedge x \notin Z) \Big) \implies x \in X \cup (Y - Z)$$

To help us figure out how to get there, let's do some scratch work here to rewrite the statement on the right-hand side; then, we can see how to get there from where we are already:

$$\begin{aligned} x \in X \cup (Y - Z) &\iff x \in X \vee x \in Y - Z && \text{Defn of } \cup \\ &\iff x \in X \vee (x \in Y \wedge x \notin Z) && \text{Defn of } - \end{aligned}$$

This is similar to the last logical equivalence we derived up above, but this one differs in the term on the left. Can you see how the one up above *implies* this one? Think about it, and then read on for the rest of our proof, resumed.

Now, we want to show that

$$\Big((x \in X \wedge x \notin Z) \vee (x \in Y \wedge x \notin Z) \Big) \implies x \in X \cup (Y - Z)$$

To do this, let's suppose the expression on the left-hand side is True. This means either

$$x \in X \wedge x \notin Z$$

or

$$x \in Y \wedge x \notin Z$$

(or possibly both). Thus, we have two cases:

1. Suppose the first expression is True, so that $x \in X \wedge x \notin Z$. This certainly means that $x \in X$, and thus $x \in X \vee x \in Y - Z$ holds.

2. Suppose the second expression is True, so that $x \in Y \wedge x \notin Z$. This means that $x \in Y - Z$, and thus $x \in X \vee x \in Y - Z$ holds.

In either case, we find that $x \in X \vee x \in Y - Z$ holds, and therefore,

$$x \in X \cup (Y - Z)$$

holds, in either case, by the definition of $\cup$.

Overall, this shows that

$$x \in (X \cup Y) - Z \implies x \in X \cup (Y - Z)$$

holds for every element $x \in U$. Therefore, by the definition of $\subseteq$, we have

$$(X \cup Y) - Z \subseteq X \cup (Y - Z)$$

□

Recognizing where we are, and where we wanted to go, helped us finish this proof. We had no hope of completing it using logical equivalences alone because, in fact, the sets given in the claim are not always equal! Looking back at the proof, can we identify the step whose logical equivalence was *invalid*, and can we use it to help construct a counterexample to the (False) claim that those sets are always equal?

We had reached as far as this valid statement

$$(x \in X \wedge x \notin Z) \vee (x \in Y \wedge x \notin Z)$$

and we used it to deduce this statement

$$x \in X \vee (x \in Y \wedge x \notin Z)$$

It seems clear, from our argument in the proof, that the first statement does *imply* the second one; that is, if we *suppose* the first statement holds, we can figure out that the second statement one holds, as well. The only difference between them is in the first term, and knowing *two* parts of an "$\wedge$" statement hold certainly lets us conclude a particular *one* of them holds.

This deduction does *not* work in the other direction. Suppose that second statement holds. If it's the right term that is valid—that $x \in Y \wedge x \notin Z$—then this also makes the first statement hold. However, since we have an "$\vee$" statement, we have to consider the case where the left term is the one that holds. In that case, knowing only $x \in X$ does not let us deduce that $x \in X \wedge x \notin Z$ holds. *Supposing* an "$\wedge$" holds lets us deduce either one of its parts holds, but just *knowing* only one part cannot tell us that both hold!

We can use this to construct a counterexample. We see that we just need to ensure that there is some particular element $x \in U$ that satisfies the left term

in the second statement, namely $x \in X$, but does *not* satisfy the left term in the first statement, namely $x \in X \wedge x \notin Z$. Said another way, we just need to ensure that there *is* an element $x \in X \cap Z$. The following example accomplishes exactly that.

Example 4.6.12. We claim that

$$(X \cup Y) - Z \subseteq X \cup (Y - Z)$$

holds for *any* sets X, Y, Z, but equaltiy *need not* hold. See the proof of Proposition 4.6.11 to see why the containment claimed above does hold.

Now, consider the following example. Let's define

$$\begin{aligned} X &= \{1\} \\ Y &= \{2\} \\ Z &= \{1, 2\} \end{aligned}$$

Notice that

$$(X \cup Y) - Z = (\{1\} \cup \{2\}) - \{1, 2\} = \{1, 2\} - \{1, 2\} = \varnothing$$

and

$$X \cup (Y - Z) = \{1\} \cup (\{2\} - \{1, 2\}) = \{1\} \cup \{\varnothing\} = \{1\}$$

Since $\varnothing \subset \{1\}$ (a *proper*) subset, we conclude that

$$(X \cup Y) - Z \neq X \cup (Y - Z)$$

in this case. This shows that equality need not hold in the claim above.

This strategy now lets us go back and complete many proofs involving sets in a more efficient and rigorous manner! Rather than fumbling through the linguistics of "ands" and "ors", we can use our logical notation and laws that we have *proven*. Many of the exercises in this section deal with sets, specifically because of this. If you need to flip back to Chapter 3 and remind yourself of any relevant definitions, go right ahead!

4.6.8 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What are the Associative Laws for Logic?

(2) What are the Distributive Laws for Logic?

(3) What are DeMorgan's Laws for Logic? What are DeMorgan's Laws for Sets? How are they related?

(4) What is the difference between a necessary and a sufficient condition?

(5) What happens when a condition is both necessary and sufficient?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) We used Truth Tables to prove DeMorgan's Laws for Logic. Can you come up with a semantic proof? Can you explain DeMorgan's Laws to a non-mathematician friend and convince them they are valid?

(2) Let $P(x)$ be the variable proposition "x is an integer that is divisible by 10". Come up with two necessary conditions and two sufficient conditions for this statement.

(3) Let A, B, C be any sets, where $A, B, C \subseteq U$, for some universal set U.

Use logical equivalences and logical laws to prove the following claims.

(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(b) $(A \cup B) \cap \overline{A} = B - A$

(c) $\overline{\overline{A} \cup B} = A \cap \overline{B}$

(d) $(A - B) \cap \overline{C} = A - (B \cup C)$

(4) Use conditional statements and logical equivalences to prove that the containment

$$A - (B \cup C) \subseteq A \cap \left(\overline{B \cap C}\right)$$

holds for any sets A, B, C.

Then, find an example that shows that equality *need not* hold.

(**Hint:** In general, a helpful idea in constructing a *strict* set containment is to see if you can make one side the *empty set*.)

(5) Let D, E, F be any sets. Consider the sets

$$D - (E - F)$$

and

$$(D - E) - F$$

How do they compare? Are they always equal? Is one always a subset of the other, or vice-versa?

Clearly state your claims, then either prove them or provide relevant counterexamples.

4.7 Negation of Any Mathematical Statement

We saw already how to negate quantified statements. With DeMorgan's Laws in hand, we now know how to negate $\wedge$ and $\vee$ statements. What's left? Aha, conditional statements!

4.7.1 Negating Conditional Statements

Consider a claim of the form $P \implies Q$. It says that *whenever* P is true, Q is also true. How do we negate such a statement? What would the logical negation even mean? Think back to how we defined " $\implies$ " as a logical connective. In which cases did we get to call the speaker of the conditional statement a *liar*. Those are precisely the cases where we would say the logical negation is True. The *only* such case was when the hypothesis P was True but the conclusion Q was False.

To prove this equivalence, we need to remember the way to write $P \implies Q$ as an "$\vee$" statement:

$$(P \implies Q) \iff (\neg P \vee Q)$$

This will help us in the proof of the following claim.

Lemma 4.7.1. *The logical negation of a conditional statement is given by*

$$\neg(P \implies Q) \iff P \wedge \neg Q$$

Proof. Observe that

$$\begin{aligned} \neg(P \implies Q) &\iff \neg(\neg P \vee Q) && \text{Logical equivalence proven already} \\ &\iff \neg(\neg P) \wedge \neg Q && \text{DeMorgan's Law for Logic} \\ &\iff P \wedge \neg Q && \text{since } \neg(\neg P) \iff P \end{aligned}$$

□

This makes intuitive sense: to show that a conditional claim is False, we need to find a case where the hypothesis holds but the conclusion fails.

Despite the risk of putting bad ideas into your head that weren't already there, we are going to point out some statements that are **NOT** logically equivalent to $\neg(P \implies Q)$. These are common mistakes that we see students use fairly often. Let's check them out and see why they don't actually work. For each of them, keep in mind that we want the logical negation $\neg(P \implies Q)$ to

be *guaranteed to have the exact opposite truth value* of the original statement $P \implies Q$. In each of these cases, we can see that this relationship would not hold.

- $\neg P \implies Q$

  This conditional statement has no logical connection to the original claim, $P \implies Q$. Remember that the statement $P \implies Q$ makes *no claim* about whether Q is true or not in the cases where P is False. (Think about the "If it is raining, then I carry my umbrella" example. If it's not raining, who knows what I'm carrying!) Thus, why would we expect Q to be *necessarily true* in those cases, like this statement says?

- $P \implies \neg Q$

  Again, this conditional statement has no logical connection to the original claim. Think about the umbrella example again. This statement would say "If it is raining, then I will **not** be carrying an umbrella." Is that what it means for the original claim to be False? Definitely not!

- $P \not\Longrightarrow Q$

  This one is more subtle. A mathematician would read "$P \not\Longrightarrow Q$" as "P does *not necessarily* imply Q". That is, it would say that there are assignments of truth values where $P \implies Q$ is valid and there are assignments where $P \implies Q$ is invalid; those cases would depend on what the individual statements P and Q are. This is a somewhat meaningful claim to make, depending on the situation, but it is not, strictly speaking, the **logical negation** of the original claim.

  In particular, we run into an issue when we try to take the logical negation of *this* statement. What does it mean to say that "It is not the case that P does not necessarily imply Q"? Does that mean there exists cases where P does not imply Q but there are also cases where P does imply Q? That sounds an awful lot like the actual claim $P \not\Longrightarrow Q$, itself ...

  For these reasons, we want to avoid using this notation: $\not\Longrightarrow$. It does have some kind of meaning in mathematics, but it is not really well-defined in a symbolic logical sense. And in any event, it is definitely *not* the logical negation of $\implies$.

Now that we have these common *errors* out of the way, let's stress the *correct* negation of $P \implies Q$. We find that it's quite helpful to remember the "$\vee$" version of a conditional statement; from there, it's easy to apply DeMorgan's Law and negate the statement:

$$\neg(P \implies Q) \iff \neg(\neg P \vee Q) \iff P \wedge \neg Q$$

Negating "$\iff$"

To negate a biconditional statement, we just write it as a conjuction of two conditional statements:

$$\neg(P \iff Q) \iff [\neg(P \implies Q) \vee \neg(Q \implies P)] \iff (P \wedge \neg Q) \vee (Q \wedge \neg P)$$

If you do any kind of comptuer programming, you might recognize the statement on the right as the `XOR` operator! It says that *exactly one* statement is True, either P or Q, but *not both*.

4.7.2 Negating Any Statement

That's it, right? We have now discussed how to negate any fundamental mathematical claim: $\exists$, $\forall$, $\wedge$, $\vee$, and $\implies$. Everything else we write will be a combination of these basic claims, so we should be able to just apply these techniques over and over and negate any statement we come across. Essentially, we just read the statement left to right and negate every piece in turn. Come across a "$\exists$"? Just switch it to a "$\forall$" and negate the property that comes after! Come across an "$\vee$"? Negate both sides and switch it to and "$\wedge$"! Come across a conditional? Apply the technique we just showed above!

Let's see several examples to get the idea.

Example 4.7.2. Find the logical negation of

$$\forall x \in \mathbb{R}\text{.}\ x < 0 \vee x > 0$$

This statement says "Every real number x satisfies either $x < 0$ or $x > 0$."

The logical negation is

$$\neg\left(\forall x \in \mathbb{R}\text{.}\ x < 0 \vee x > 0\right) \iff \exists x \in \mathbb{R}\text{.}\ x \geq 0 \wedge x \leq 0$$

Notice that we applied DeMorgan's Law for Logic to negate the $\vee$ claim on the right-hand side, and we used the fact that $x \not> 0$ is logically equivalent to $x \leq 0$.

We see that this negation is True because $0 \in \mathbb{R}$ and $0 \leq 0$ and $0 \geq 0$. Thus, the original statement was False.

Example 4.7.3. Find the logical negation of

$$\exists n \in \mathbb{N}\text{.}\ n \geq 10 \wedge \sqrt{n} \leq 3$$

This statement says "There exists a natural number n that satisfies both $n \geq 10$ and $\sqrt{n} \leq 3$.

The logical negation is

$$\forall n \in \mathbb{N}\text{.}\ n < 10 \vee \sqrt{n} > 3$$

That is, the logical negation says "Every natural number n satisfies either $n > 10$ or $\sqrt{n} < 3$.

Example 4.7.4. Find the logical negation of

$$\exists x \in \mathbb{R}\,.\ \forall y \in \mathbb{R}\,.\ x \geq y \implies x^2 \geq y^2$$

This statement says "There exists a real number x such that whenever we have a real number y that satisfies $x \geq y$, we may conclude that $x^2 \geq y^2$".

The logical negation is

$$\forall x \in \mathbb{R}\,.\ \exists y \in \mathbb{R}\,.\ x \geq y \wedge x^2 < y^2$$

Can you prove that this logical negation is, in fact, the True statement? Try it!

Example 4.7.5. Find the logical negation of

$$\forall X \in \mathcal{P}(\mathbb{Z})\,.\ (\forall x \in X\,.\ x \geq 1) \implies X \subseteq \mathbb{N}$$

This statement says that "For every subset X of the integers $\mathbb{Z}$, if every element x of the set X satisfies $x \geq 1$, then X is a subset of the natural numbers $\mathbb{N}$."

The logical negation is

$$\exists X \in \mathcal{P}(\mathbb{Z})\,.\ (\forall x \in X\,.\ x \geq 1) \wedge X \not\subseteq \mathbb{N}$$

This statement says that "There is a subset $X \subseteq \mathbb{Z}$ such that every element $x \in X$ satisfies $x \geq 1$ and yet $X \not\subseteq \mathbb{N}$." We could even rewrite the last part further by noting that

$$X \not\subseteq \mathbb{N} \iff \exists y \in X\,.\ y \notin \mathbb{N}$$

although this wouldn't be totally essential.

Which statement (the original or the negation) is True? Can you prove it?

Compare the statement used in the example above with the following one:

$$\forall X \in \mathcal{P}(\mathbb{Z})\,.\ \forall x \in X\,.\ \big(x \geq 1 \implies X \subseteq \mathbb{N}\big)$$

The only difference between them is the location of the parentheses, but this completely changes the statement's meaning!

The statement used in the example asserts something about *every* subset of the integers. That is, no matter what subset $X \subseteq \mathbb{Z}$ is introduced, the statement says that *if* that set has the property that all of its elements are at least 1, *then* that set X is actually a subset of $\mathbb{N}$, as well.

The new statement written in this box says something else: no matter what subset $X \subseteq \mathbb{Z}$ is introduced and, furthermore, no matter what element

x of *that* set X is introduced, the statement says that *if* that element x is at least 1, *then* that set X is a subset of $\mathbb{N}$, as well.

Do you see why this is different? The issue is where the "if" happens: where does the quantification end and the conditional statement begin? The first statement, from the above example, puts the quantification over elements of X inside the "if" part of the conditional statement. The second statement, in this box, puts that quantification before the conditional statement entirely.

We claim that this second version, in this box, is False, and we encourage you to figure out why (if you haven't already).

Example 4.7.6. Let $O(x)$ be the proposition "x is odd", and let $E(x)$ be the proposition "x is even". Find the logical negation of the statement

$$\forall x, y \in \mathbb{Z}\text{.}\ O(x \cdot y) \iff (O(x) \wedge O(y))$$

This statement says that "For every two integers x and y, their product is odd if and only if they are both odd, themselves".

Before we find the logical negation, remember the $\iff$ means "$\implies$ and $\impliedby$". Let's rewrite the claim that way first, so that we can negate it properly:

$$\forall x, y \in \mathbb{Z}\text{.}\ [O(x \cdot y) \implies (O(x) \wedge O(y))] \wedge [(O(x) \wedge O(y)) \implies O(x \cdot y)]$$

The logical negation is

$$\begin{aligned}
&\neg\Big(\forall x, y \in \mathbb{Z}\text{.}\ [O(x \cdot y) \implies (O(x) \wedge O(y))] \wedge [(O(x) \wedge O(y)) \implies O(x \cdot y)]\Big) \\
&\quad\iff \exists x, y \in \mathbb{Z}\text{.}\ \neg\,[O(x \cdot y) \implies (O(x) \wedge O(y))] \\
&\qquad\qquad\qquad \vee \neg\,[(O(x) \wedge O(y)) \implies O(x \cdot y)] \\
&\quad\iff \exists x, y \in \mathbb{Z}\text{.}\ [O(x \cdot y) \wedge \neg(O(x) \wedge O(y))] \vee [(O(x) \wedge O(y)) \wedge \neg O(x \cdot y)] \\
&\quad\iff \exists x, y \in \mathbb{Z}\text{.}\ [O(x \cdot y) \wedge (E(x) \vee E(y))] \vee [(O(x) \wedge O(y)) \wedge E(x \cdot y)]
\end{aligned}$$

That is, the logical negation says "There exist integers x and y such that either their product is odd and yet (at least) one of them is even, or they are both odd and yet their product is even.

Can you prove which one of these claims is True?

4.7.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How is a mathematical statement related to its logical negation?

(2) What is the logical negation of a conditional statement?

(3) Consider the statement $P \implies Q$. What is its contrapositive? What is the logical negation of that contrapositive? Can you see that it must have the *same* truth value as the logical negation of the original statement?

(4) What is the logical negation of an *if and only if* statement, $P \iff Q$? Why does this make sense, considering what such a statement says about the *truth values* of P and Q?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Write out the logical negation of each of the following mathematical statements.

Then, determine the *truth value* of each statement.

(If you're feeling ambitious, formally prove/disprove each statement!)

(a) $\exists x \in \mathbb{N}\text{.}\ \forall y \in \mathbb{N}\text{.}\ y - x^2 \geq 0$

(b) $\exists x \in \mathbb{Z}\text{.}\ \forall y \in \mathbb{R}\text{.}\ xy = 0$

(c) $\exists x \in \mathbb{Z}\text{.}\ \forall y \in \mathbb{Z}\text{.}\ (y \neq 0 \implies xy > 0)$

(d) $\forall a, b \in \mathbb{Q}\text{.}\ ab \in \mathbb{Z} \implies (a \in \mathbb{Z} \vee b \in \mathbb{Z})$

(e) $\forall x \in \mathbb{R}\text{.}\ x > 0 \implies \big(\exists y \in \mathbb{R}\text{.}\ y < 0 \wedge xy > 0\big)$

(f) $\forall x \in \mathbb{R}\text{.}\ \left|x + \frac{1}{x}\right| = 2 \iff x = 1$

(2) Let $A = \{1, 2, 3, 4\}$ and $B = \{2, 3\}$.

What is the difference between the following two statements? Determine the truth value of each one.

Then, negate each one, and explain how those negations also differ. What are their truth values?

(a) $\forall x \in A\text{.}\ \forall y \in B\text{.}\ (x \geq y \implies x^2 \geq 4)$

(b) $\forall x \in A\text{.}\ (\forall y \in B\text{.}\ x \geq y) \implies x^2 \geq 4$

(3) Let $P = \{x \in \mathbb{R} \mid x > 0\}$. Write the logical negation of each of the following statements, and determine their truth values.

(a) $\forall \varepsilon \in P\text{.}\ \forall x \in P\text{.}\ \exists \delta \in P\text{.}\ \forall y \in \mathbb{R}\text{.}\ \Big(|x - y| < \delta \implies \left|\frac{1}{x} - \frac{1}{y}\right| < \varepsilon\Big)$

(b) $\forall \varepsilon \in P\,.\, \exists \delta \in P\,.\, \forall x \in P\,.\, \forall y \in \mathbb{R}\,.\, \Big(|x-y| < \delta \implies \left|\frac{1}{x} - \frac{1}{y}\right| < \varepsilon \Big)$

Hint/suggestion: A statement like $|a| < b$ can be written as $-b < a < b$. Also, a statement like $a < b < c$ can be written as $a < b \wedge b < c$. This might help you rewrite the statements when determining their truth values.

(4) Let $P(n)$ be the proposition "n is odd" and let $Q(n)$ be the proposition "$n^2 - 1$ is divisible by 8".

Write the logical negation of the statement

$$\forall n \in \mathbb{N}\,.\, P(n) \iff Q(n)$$

and determine its truth value.

(5) Let $P = \{x \in \mathbb{R} \mid x > 0\}$. Write the logical negation of each of the following statements, and determine their truth values.

(a) $\forall \varepsilon \in P\,.\, \exists \delta \in P\,.\, \forall x \in \mathbb{R}\,.\, 0 < x < \delta \implies \frac{1}{x} > \frac{10}{\varepsilon}$

(b) $\forall \varepsilon \in P\,.\, \exists x \in \mathbb{R}\,.\, \forall n \in \mathbb{N}\,.\, \big(n > x \implies \frac{(-1)^n}{n} < \varepsilon \big)$

(c) $\forall \varepsilon \in P\,.\, \exists x \in \mathbb{R}\,.\, \forall n \in \mathbb{N}\,.\, \big(n > x \iff \frac{(-1)^n}{n} < \varepsilon \big)$

4.8 [Optional Reading] Truth Values and Sets

There is a convenient and interesting relationship between sets (and their corresponding relationships and operations) and logical truth values (and their corresponding relationships and connectives). We will mention it here and demonstrate some examples, and leave it to you to investigate it further, if you'd like. We won't need these kinds of ideas in the rest of our work, but we believe that thinking about these ideas and sorting them out in your head will really help your understanding of the fundamentals of logic, as well as sets.

Suppose we have two variable propositions, $P(x)$ and $Q(x)$. Further, suppose these propositions make sense for any input x that comes from some universal set U. (This set U depends on the specific statements inside $P(x)$ and $Q(x)$, of course, but we don't really care what they are for this general discussion.) For each of these propositions, we can define a **truth set**; that is, we can consider the set of all instances x from the universe U that make those propositions evaluate as True. We define

$$T_P = \{x \in U \mid P(x) \text{ is } \mathsf{True}\}$$
$$T_Q = \{x \in U \mid Q(x) \text{ is } \mathsf{True}\}$$

Perhaps the propositions $P(x)$ and $Q(x)$ are related somehow. Let's suppose that, in fact,

$$\forall x \in U\,.\, P(x) \implies Q(x)$$

holds. What does this say about those **truth sets**? This conditional statement says that any x that satisfies $P(x)$ (i.e. makes $P(x)$ True) must *also* satisfy $Q(x)$. Written another way, using those truth sets, we have

$$\forall x \in U \,.\, x \in T_P \implies x \in T_Q$$

This is precisely the definition of "subset"! What we have just discovered is that

$$T_P \subseteq T_Q$$

when that conditional statement above holds.

Let's suppose even further, now, that

$$\forall x \in U \,.\, P(x) \iff Q(x)$$

holds. We just discovered that $T_P \subseteq T_Q$, and applying the exact same reasoning to the "other direction" of that $\iff$ statement (that is, the $Q(x) \implies P(x)$ part of it) will show us that $T_Q \subseteq T_P$, as well. By the definition of set equality, this means that

$$T_P = T_Q$$

when that biconditional statement above holds.

How else might we combine the propositions $P(x)$ and $Q(x)$? Let's consider the proposition $P(x) \wedge Q(x)$. What are the instances x that make this conjunction True? How can we characterize those instances in terms of the truth sets we defined? Think about it for a minute, and you'll see that all of those instances are characterized by the intersection of those sets; we need *both* $P(x)$ and $Q(x)$ to hold, so we need an instance that comes from both of the truth sets.

Similarly, we can consider the conjunction $P(x) \vee Q(x)$. An instance x makes this statement True when that x makes *at least one* of the propositions True. Thus, that x must come from at least one of the sets, so it must come from their *union*.

Let's summarize these relationships we have discovered:

$$\begin{aligned}
&\forall x \in U \,.\, \big(\, P(x) \implies Q(x) \,\big) \iff \big(\, T_P \subseteq T_Q \,\big) \\
&\forall x \in U \,.\, \big(\, P(x) \iff Q(x) \,\big) \iff \big(\, T_P = T_Q \,\big) \\
&\forall x \in U \,.\, \big(\, P(x) \wedge Q(x) \,\big) \iff \big(\, x \in T_P \cap T_Q \,\big) \\
&\forall x \in U \,.\, \big(\, P(x) \vee Q(x) \,\big) \iff \big(\, x \in T_P \cup T_Q \,\big)
\end{aligned}$$

Can you come up with some characterizations, using truth sets, for the following statements? Fill in the blanks!

$$\forall x \in U \,.\, \big(\, P(x) \wedge \neg Q(x) \,\big) \iff \underline{\qquad\qquad}$$

$$\big(\, \exists x \in U \,.\, P(x) \,\big) \iff \underline{\qquad\qquad}$$

$$\forall x \in U \,.\, \big(\, \neg P(x) \iff \underline{\qquad\qquad} \,\big)$$

(Careful: how are the previous statement and the next one different?)

$$\big(\, \forall x \in U \,.\, \neg P(x) \,\big) \iff \underline{\qquad\qquad}$$

4.9 Writing Proofs: Strategies and Examples

We are now prepared to fully tackle the goal we have been building towards all along: writing PROOFS!

In this section, we will apply all of these fundamental logical principles we have developed in this chapter. Specifically, we will learn how to use them to write formal arguments that demonstrate the truth (or falsity) of mathematical statements. In general, it's hard to describe how to figure out which mathematical statements are True and which are False. In a way, though, the strategies we develop here will help us discover truths. More importantly, though, they will provide us with templates and guidelines for how to actually present a truth to someone else and describe *why* it is, indeed, a truth.

As we have discussed, it is not enough to figure out some interesting fact and just hope that others will trust us about it. We need to be able to *explain* that fact; we need to present an argument that will *convince* someone else of its truth. We don't necessarily have to explain where it came from, or why we cared to investigate it in the first place (although sometimes you might want to answer these implicit questions, if you think it would help a potential reader). In general, we just need to make sure that someone else—a peer, a classmate, a fellow mathematician—can pick up our written proof, read it, and afterwards be fully convinced that what we claimed to be True is, indeed, True.

Outline of this section

Mostly, we hope you will see how the ensuing strategies come directly from the underlying logical principles associated with propositions and quantifiers and connectives and negations. We have split the section into several subsections, each one corresponding to a particular quantifier or connective.

When you face a mathematical statement and have to prove it, just start reading the statement left to right. What do you encounter first? If it's a "$\exists$" quantifier, look at Section 4.9.1. If it's a "$\forall$" quantifier, look at Section 4.9.2. After that, what type of claim do you face? What form does the ensuing variable proposition take? Is it an "$\vee$" statement? Look at Section 4.9.3. Is it a conditional statement? Look at section 4.9.5. Is it a conditional statement where the hypothesis is an "$\vee$" statement and the conclusion is an "$\wedge$" statement? Look at all three Sections—4.9.3 and 4.9.4 and 4.9.5—and combine them appropriately! In general, every proof we write from now on (except for induction proofs, which we will return to in the next chapter) will be a combination of these strategies. Which ones you use and how you combine them depends on the statement you're trying to prove and how you've decided to approach it.

Within each subsection, we have provided some templates and some examples. You might find the templates too restrictive, perhaps too formal; we understand, but we think that following our formats as closely as possible for now will help you in the long run. These templates—as well as how we've used them in the examples provided—are meant to emphasize the logical principles behind these proof strategies. Working with them closely will give you extra

practice with these logical concepts and, we strongly believe, help you adapt them for your own uses in the future.

For each example provided, we have boxed the proof strategy in blue and the example implementation in green and any necessary scratch work in red. Any other discussion of the strategy or the implementation appears outside of those boxes.

Also, several of the examples we consider in this section (and the next one) are interesting and useful results, in their own right. You'll notice that some of them have a name or a descriptive title, which is meant to indicate this fact. While the main emphasis of this section is on the **proof strategies**—developing them and seeing how to use them—we encourage you to also keep in mind these examples as interesting facts, themselves. We'll bring up this idea again when it's warranted, but we'll keep those discussions brief, so as not to distract from the overall structure of this section.

Direct vs. Indirect methods

You will also notice that each subsection includes strategies for both **direct** and **indirect** methods. These terms might not be familiar to you yet. All they refer to is whether or not we try to prove a statement (1) directly by demonstrating that it is True, or (2) indirectly by invoking the Law of the Excluded Middle, by demonstrating that its logical negation is False.

Both forms of strategy are, in general, equally valid, but **direct** proofs are typically considered subjectively better by many readers. (Sometimes, you might write an indirect proof that is actually hiding a direct proof inside it!) These subjective ideas will be assessed and discussed as we work through examples and ask you to write proofs on your own, in the exercises.

You'll notice that all of our indirect proofs begin with the phrase "Assume for sake of contradiction", usually abbreviated as "AFSOC". This is an important and helpful phrase. It signals to the reader of our proof that we are going to make an assumption but we don't *really* think that the assumption is valid. Rather, we are going to use this assumption to derive something False, a **contradiction**. This will allow us to conclude that our original assumption was invalid, so its logical negation (i.e. our original statement that we hoped to prove) is actually True. You'll see that we use the symbol " $\lightning$ " to indicate a contradiction, but we also make sure to point out *why* we have found a contradiction. We don't leave it to the reader to figure it out!

Alright, that's enough preamble. Let's dive right in and WRITE SOME PROOFS!

4.9.1 Proving $\exists$ Claims

An "$\exists$" claim is one of *existence*. It asserts that some particular object exists as an element of some set and that it has a certain property. To prove such a claim, we need to exhibit such an object and verify, for our reader, that (1)

that object is an element of the correct set and (2) that object has the correct property.

Direct Method

Strategy:

Claim: $\exists x \in S \text{ . } P(x)$

Direct proof strategy:

Define a specific example, $y =$ _____.

Prove that $y \in S$.

Prove that $P(y)$ holds true.

Example 4.9.1. *Solving a system of linear equations:*

Statement: Fix $a, b, c, d, e, f \in \mathbb{R}$ with the property that $ad - bc \neq 0$.

We claim that one can simultaneously solve

$$ax + by = e \tag{4.1}$$
$$cx + dy = f \tag{4.2}$$

for some $x, y \in \mathbb{R}$.

Define $S(x, y)$ to be the statement "x and y simultaenously satisfy both equations, (4.1) and (4.2), above". Then we claim

$$\exists x, y \in \mathbb{R} \text{ . } S(x, y)$$

First, we must do some scratch work to *construct* the solution. Then, we can write a proof that defines the objects x and y and shows why they work.

Scratch work:

We need $ax + by = e$ and $cx + dy = f$, and we want to know which x and y make this work.

Let's multiply the first and second equations by the right coefficients (namely, d and -b, respectively) so we can cancel the y terms by adding

the two lines:

$$\begin{aligned} adx + bdy &= de \\ +\big(-bcx - bdy &= -bf\big) \\ \hline (ad-bc)x &= de - bf \end{aligned}$$

Dividing tells us $x = \frac{de-bf}{ad-bc}$, which is okay because $ad - bc \neq 0$.

Doing something similar, canceling the x terms, tells us how to get y:

$$\begin{aligned} acx + bcy &= ce \\ +\big(-acx - ady &= -af\big) \\ \hline (bc-ad)y &= ce - af \end{aligned}$$

Dividing tells us $y = \frac{af-ce}{ad-bc}$.

The main lesson here is that we do not need to show this scratch work in our proof below! We don't assume that a reader would care to wade through our messy notes about *how* we came up with the solution to the system of linear equations. Rather, we assume the the reader only cares about *what* the solution is and *why* it's a solution. Also, this makes the proof much more concise, so it can be read more easily and quickly.

Implementation:

Proof. Since $ad - bc \neq 0$ (by assumption), we may define

$$x = \frac{de-bf}{ad-bc} \qquad \text{and} \qquad y = \frac{af-ce}{ad-bc}$$

and know that $x, y \in \mathbb{R}$. Then,

$$ax + by = \frac{(ade - abf) + (abf - bce)}{ad-bc} = \frac{ade - bce}{ad-bc} = \frac{e(ad-bc)}{ad-bc} = e$$

$$cx + dy = \frac{(cde - bcf) + (adf - cde)}{ad-bc} = \frac{adf - bcd}{ad-bc} = \frac{f(ad-bc)}{ad-bc} = f$$

so $S(x, y)$ holds. □

If you've studied some linear algebra before, you'll recognize the term $ad-bc$ as the **determinant** of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The stipulation that $ad - bc \neq 0$ means that we require this matrix of coefficients to *have an inverse*, to be "non-

singular". In that situation, we have a solution to the system for any $e, f \in \mathbb{R}$.

Indirect Method (Proof by Contradiction)

This strategy relies on the logical negation of an $\exists$ claim:

$$\neg\big(\exists x \in S\,.\; P(x)\big) \iff \forall x \in S\,.\; \neg P(x)$$

We will assume this negation and deduce something contradictory from it, meaning the negation is False so the original is True.

Claim: $\exists x \in S\,.\; P(x)$

Indirect proof strategy:

AFSOC that for every $y \in S$, $\neg P(y)$ holds.

Find a contradiction.

Example 4.9.2. *A version of the Pigeonhole Principle:*

Statement: Suppose $n \in \mathbb{N}$ and we have n real numbers, $a_1, a_2, \ldots, a_n \in \mathbb{R}$.

We claim that one of the numbers is at least as large as their average. That is,

$$\exists B \in [n]\,.\; a_B \geq \frac{1}{n}\sum_{i=1}^{n} a_i$$

Proof. AFSOC none of the numbers are at least as large as the average, i.e.

$$\forall i \in [n]\,.\; a_i < \frac{1}{n}\sum_{i=1}^{n} a_i$$

Define the constant $S = \sum_{i=1}^{n} a_i$, so that $a_i < \frac{S}{n}$.

Then we can sum all of the a_is and observe that

$$S = \sum_{i=1}^{n} a_i < \sum_{i=1}^{n} \frac{S}{n} = n \cdot \frac{S}{n} = S$$

This shows that the real number S is *strictly* less than itself: $S < S$. This is a contradiction. ⨳

Therefore, the original assumption was false, and the claim follows. □

As stated, this is a version of the **Pigeonhole Principle**. We will investigate and use this principle again in Section 8.6, when we study **combinatorics**.

4.9.2 Proving ∀ Claims

A "∀" claim is one of *universality*. It asserts that *all* elements of a set have some common property. To prove such a claim, we need to show that *every* element of the set has that property. To accomplish this "all at once", we will consider an **arbitrary and fixed** element of the set, and prove that it has the desired property. Because this object is arbitary, our argument applies to every element of the set. Because this object is fixed, we are allowed to refer to it by name throughout the proof.

Direct Method

Strategy:

Claim: $\forall x \in S\,.\, P(x)$

Direct proof strategy:

Let $y \in S$ be arbitrary and fixed.

Prove that $P(y)$ holds true.

Example 4.9.3. *A version of the AGM Inequality:*

Statement: $\forall x, y \in \mathbb{R}\,.\, xy \leq \left(\frac{x+y}{2}\right)^2$

Implementation:

Proof. Let $x, y \in \mathbb{R}$ be arbitrary and fixed.

We know $0 \leq (x-y)^2$.

Multiplying out and rearranging, we get $2xy \leq x^2 + y^2$.

Adding $2xy$ to both sides, we get $4xy \leq x^2 + 2xy + y^2$.

Factoring, we get $4xy \leq (x+y)^2$.

Dividing by 4 and putting that into the square, we get

$$xy \leq \left(\frac{x+y}{2}\right)^2$$

□

This result is known as the **AGM Inequality** because it deals with the Arithmetic Mean (AM) and the Geometric Mean (GM) of two real numbers.

The Arithmetic Mean of x and y is $\frac{x+y}{2}$.

The Geometric Mean of x and y is $\sqrt{xy}$. (Notice that this only applies when

$xy \geq 0$, i.e. when x and y have the *same sign* be it positive or negative, or zero.)

The AGM Inequality asserts that the AM is always at least as large as the GM. A helpful mnemonic is to read "**AGM**" as "**A**rithmetic Mean **G**reater than **G**eometric **M**ean".

What we proved above is a slightly more general version, because it applies to *all* real numbers x and y, and not just those with the same sign. Supposing that $xy \geq 0$, though, one can simply take the square root of both sides and obtain the "usual" statement of the AGM Inequality: $\sqrt{xy} \leq \frac{x+y}{2}$.

Indirect Method (Proof by Contradiction)

Claim: $\forall x \in S \text{ . } P(x)$

Indirect proof strategy:

AFSOC that $\exists y \in S$ such that $\neg P(y)$ holds.

Find a contradiction.

Example 4.9.4. *$\sqrt{2}$ is irrational:*

Statement: $\forall a, b \in \mathbb{Z} \text{ . } \frac{a}{b} \neq \sqrt{2}$

(Note: This claim is appealing directly to the definition of the rational numbers $\mathbb{Q}$. It is saying that $\sqrt{2} \notin \mathbb{Q}$ because that number has *no* representation as a ratio of integers.)

Implementation:

Proof. AFSOC $\exists a, b \in \mathbb{Z} \text{ . } \frac{a}{b} = \sqrt{2}$.

We may assume that $\frac{a}{b}$ is given in *reduced form*, so that a and b have no common factors. (If this were not the case, we could just reduce the fraction and obtain such a form.)

Let such $a, b \in \mathbb{Z}$ be given.
(**Note:** We discuss this phrase, "let such _____ be given", below in Section 4.9.8. It is meant to not only assert that such an $a, b \in \mathbb{Z}$ *exist*, but also that we want some *particular* instances of those variables so that we can work with them for the rest of the proof.)

This means $\frac{a}{b} = \sqrt{2}$, so $\frac{a^2}{b^2} = 2$.

Thus, $2b^2 = a^2$, so a^2 is even, by definition.

Since a^2 is even, this tells us a is even.

(**Note:** You should prove this. We will prove it in Section 4.9.6, but you should try to do it on your own now.)

Thus, $\exists k \in \mathbb{Z}.\ a = 2k$. Let such a k be given, so $a^2 = 4k^2$.

Then $2b^2 = 4k^2$, so $b^2 = 2k^2$.

Thus, b^2 is even, by definition. By the same reasoning as above with a^2 and a, we deduce here that b is even.

This shows both a and b are even so, in fact, they have a common factor of 2.

This contradicts our assumption that a and b have no common factors. ※

Therefore, the original assumption must be False, so the claim is True. □

4.9.3 Proving $\vee$ Claims

An "$\vee$" claim asserts that at least one of two statements must be True. If it just so happens that one of the two statements is clearly False, then just try to prove the other one is True. This is what the direct method is here; it is straightforward, so we will not provide an example of implementation.

Direct Method

Strategy:

Claim: $P \vee Q$

Direct proof:

Prove that P is True, or else prove that Q is True.

This relies on you being able to decide ahead of time which one of the statements (P or Q) is True, of course. If you can do so, then this isn't even really a "strategy". Just implement whatever strategy applies to P (or Q, as the case may be).

Indirect Method (Proof by "Otherwise")

This method is far more interesting than the direct one. In general, it is helpful when the statements P and Q are actually variable propositions, and for some instances P is True whereas for other instances Q is the True one. In that case, rather than characterize *exactly* which instances satisfy P and which satisfy Q, we can just say, "Well, if P is True, then our proof is already complete. Thus, all we need to worry about are the cases where P is False; for those cases, we need to show that Q is still True."

Strategy:

Claim: $P \vee Q$

Indirect proof strategy 1:

Suppose that $\neg P$ holds. Prove that Q holds.

To reiterate the strategy: If P holds, then the claim holds, and in fact we don't care whether Q holds or not. *Otherwise*, in the case where P fails, we need to guarantee that Q holds.

Notice that $P \vee Q \iff Q \vee P$ (i.e. the order is irrelevant in a logical disjunction) so we can rewrite our claim as $Q \vee P$ and rewrite the above strategy as:

Suppose $\neg Q$ holds. Prove that P holds.

This is the exact same strategy on the equivalent statement $Q \vee P$, i.e. with the roles of P and Q switched.

Example 4.9.5. *When a real number is less than its square:*

Statement: Suppose that $x \in \mathbb{R}$ and $x^2 \geq x$.

We claim that $x \geq 1$ or $x \leq 0$.

Implementation:

Proof. Let $x \in \mathbb{R}$ be arbitrary and fixed, and suppose that $x^2 \geq x$.

If it were the case that $x \leq 0$, we would be done; so, suppose otherwise.

That is, suppose $x > 0$.

By assumption, $x^2 \geq x$. Since $x > 0$, we can divide both sides by x.

This yields $x \geq 1$. □

This proves a *necessary* condition for a real number to be less than (or equal to) its square. Is this condition—namely, $x \geq 1 \vee x \leq 0$—also a *sufficient* condition? Prove it! It's easy, and once you've done so, we will have together proven this *biconditional* statement:

$$\forall x \in \mathbb{R}\text{.}\ x^2 \geq x \iff (x \geq 1 \vee x \leq 0)$$

Indirect Method (Proof by Contradiction)

This method is more like the indirect methods considered above, in that we suppose the logical negation is valid and then deduce something absurd. We will illustrate this strategy by applying it to the same claim in the directly previous example.

Strategy:

Claim: $P \vee Q$

Indirect proof strategy 2:

AFSOC that $\neg P \wedge \neg Q$ holds. Find a contradiction.

Implementation:

Proof. Let $x \in \mathbb{R}$ be arbitrary and fixed, and suppose that $x^2 \geq x$.

AFSOC that $0 < x$ and $x < 1$.

Since $x > 0$, we can multiply both sides of these inequalities by x and preserve the sign.

Multiplying $x < 1$ by x, then, we find that $x^2 < x$.

This contradicts our assumption that $x^2 \geq x$ ※

Thus, our assumption was invalid, so the claim is True. □

How does this compare to the previous implementation? We are proving the exact same claim, but the proofs are slightly different. Which do you think is better, in your opinion? Which do you think was easier to write? Furthermore, could you go back and rewrite the original claim using quantifiers and "$\implies$"? After doing that, do you see what these two proofs accomplish? Try it!

4.9.4 Proving $\wedge$ Claims

An "$\wedge$" claim asserts that both of two statements are True. There's one obvious and direct way to do this: just prove one statement and then prove the other!

We will show you an example implementation of this method because the $\wedge$ statement of our example comes *after* an $\exists$ claim. Thus, there's actually some scratch work to be done to figure out how to define an object that will, indeed, satisfy both of the desired properties. This will be our first illustrative example of how these proof strategies can be **combined** to prove statements that use both quantifiers and connectives.

Direct Method

Strategy:

Claim: $P \wedge Q$

Direct proof strategy:

Prove that P holds. Prove that Q holds.

Example 4.9.6. **A smaller number whose square is bigger:**

Statement: $\forall x \in \mathbb{R}.\ \exists y \in \mathbb{R}.\ \big(\, x \geq y \wedge x^2 < y^2 \,\big)$

Scratch work:

How does this work? Let's take a specific x, like $x = 4$. We need to find a smaller real number whose square is bigger than $x^2 = 16$.

The key is that we want $y \in \mathbb{R}$, so we can use *negative* numbers. In this case, picking a negative number with larger *magnitude*, like $y = -5$, will work.

Let's take a different x, like $x = -2$. This number is already negative, so just picking any smaller number, like $y = -3$, will work.

The proof we follow below splits into two cases, based on whether x is positive or non-positive.

Now we are ready to prove our claim.

Implementation:

Proof 1. Let $x \in \mathbb{R}$ be arbitrary and fixed. We consider two cases.

(1) Suppose $x \leq 0$.

Define $y = x - 1$. Notice $y \in \mathbb{R}$.

Notice $y \leq x$. Also, notice that

$$y^2 = (x-1)^2 = x^2 - 2x + 1 = x^2 - (2x - 1)$$

Since $x \leq 0$, we know $2x \leq 0$ and so $2x - 1 \leq -1$. Thus,

$$x^2 - (2x - 1) \geq x^2 - 1 > x^2$$

and therefore, $y^2 > x^2$.

(2) Now, suppose $x > 0$.

Define $y = -x - 1$. Notice $y \in \mathbb{R}$.

Notice $y < 0$ and $x > 0$, so $y \leq x$. (In fact, $y < x$, even.)

Also, notice that

$$y^2 = (-x-1)^2 = x^2 + 2x + 1 = x^2 + (2x + 1)$$

Since $x > 0$, we know $2x + 1 > 0$. Thus,

$$x^2 + (2x + 1) > x^2$$

and therefore, $y^2 > x^2$.

In either case, we found a y with the desired properties, namely $y \in \mathbb{R}$ and $y \leq x$ and $x^2 < y^2$. Therefore, the claim is True. □

Why did we call this "Proof 1"? We split the proof into two cases based on our observations in the scratch work. Specifically, we recognized that we will define y (in terms of x) *differently*, depending on the sign of x. We claim that it is possible to rewrite this proof in a way that *avoids* these cases. This is what "Proof 2" will be, and we want you to write it! To reiterate the goal, we want you to rewrite the above proof so that y is defined in terms of x in one general way that works regardless of the sign of x.

(*Hint:* What is $-x$ when $x < 0$? Is this a function we've seen before?)

Indirect Method (Proof by Contradiction)

This method is just like the other indirect methods. We just take the logical negation of a claim, assume it holds, and deduce something absurd. This means that the assumption was invalid, so the original statement is the True one.

We will leave it to you to try to apply this method to the claim used in the previous example. (Note: You might want to do this *after* finding the "second proof" we hinted at just above this.) Then, you can compare how the two methods played out and decide which one you prefer, in this case.

Strategy:

Claim: $P \wedge Q$

Indirect proof:

AFSOC that $\neg P \vee \neg Q$ holds.

Consider the first case, where $\neg P$ holds. Find a contradiction.

Consider the second case, where $\neg Q$ holds. Find a contradiction.

4.9.5 Proving $\Longrightarrow$ Claims

It might help you to look back at Section 4.5.3, where we introduced the connective "$\Longrightarrow$". Specifically, we want you to recall that $P \implies Q$ means that *whenever* P holds, Q also *necessarily* holds. This conditional statement is True

in the cases where P itself (the **hypothesis**) is False. Thus, our proof strategy does not need to consider such cases. All we need to do is *suppose* that P holds, and deduce that Q also holds. This takes care of the "whenever P holds, so does Q" consideration.

Direct Method

Strategy:

Claim: $P \implies Q$

Direct proof strategy:

Suppose P holds. Prove that Q holds.

Example 4.9.7. *Monotonicity of squares:*

Statement: $\forall y \in \mathbb{R}.\ y > 1 \implies y^2 - 1 > 0$

Implementation:

Proof. Let $y \in \mathbb{R}$ be arbitrary and fixed.

Suppose $y > 1$.

Multiplying both sides by y (since $y > 0$), we obtain $y^2 > y$.

Since $y > 1$, this tells us $y^2 > y > 1$, and so $y^2 > 1$.

Subtracting yields the desired conclusion: $y^2 - 1 > 0$. □

We called this "monotonicity of squares" because it states a particular property of real numbers that is **monotone**. This is a term that is used to indicate a certain inequality is preserved under an operation. In this case, the fact that some number being greater than 1 is preserved by the "squaring operation". That is, we proved that if $y > 1$, then $y^2 > 1^2$, as well.

Now, this was a pretty easy example to prove, but we wanted to include it to emphase the proof strategy for conditional statements. Let's work with a more difficult example now.

(You'll also notice that Exercise 4.11.22 has a similar-looking problem statement. Perhaps you want to work on that one after following this example.)

Example 4.9.8. *Working with inequalities:*

Statement: We define the following variable propositions:

$$P(x) \text{ is } \text{“}\frac{x-3}{x+2} > 1 - \frac{1}{x}\text{”}$$
$$Q(x) \text{ is } \text{“}\frac{x+3}{x+2} < 1 + \frac{1}{x}\text{”}$$

Define $S = \{x \in \mathbb{R} \mid x > 0\}$.

We claim that

$$\forall x \in S \text{ . } P(x) \implies Q(x)$$

Scratch work:

We're guessing that a direct method will work here, so let's try to manipulate the inequality stated inside $P(x)$ and make it "look like" the inequality inside $Q(x)$.

So we start with that inequality

$$\frac{x-3}{x+2} > 1 - \frac{1}{x}$$

and we'll try multiplying both sides by $x+2$. Can we do this? Oh right, $x > 0$ and so $x + 2 > 0$, as well. Phew! This gives us

$$x - 3 > (x+2) - \frac{x+2}{x} = x + 2 - 1 - \frac{2}{x} = x + 1 - \frac{2}{x}$$

We want to see an $x + 3$ somewhere, so we'll add/subtract on both sides:

$$x - 1 + \frac{2}{x} > x + 3$$

Can we divide by $x + 2$ and make the right fraction? Hmm . . . Oh wait! We already simplified the fraction $\frac{x+2}{x}$ and moved it to one side. Maybe we shouldn't have simplified it first, so let's try undoing that:

$$x + 3 < x - 1 + \frac{2}{x} = (x+2) + \frac{x+2}{x} - 4 = (x+2)\left(1 + \frac{1}{x}\right) - 4$$

Aha, that looks better! We even have some "wiggle room" in the form of the negative 4 there. We know the right-hand side is less than what we wanted it to be, so the result holds.

Let's take these algebraic steps we worked on here, reorder them a bit and explain them, and wrap it all together in a formal proof.

Implementation:

Proof. Let $x \in S$ be arbitrary and fixed.

Suppose that $P(x)$ holds; that is, suppose

$$\frac{x-3}{x+2} > 1 - \frac{1}{x}$$

We will show that the inequality

$$\frac{x+3}{x+2} < 1 + \frac{1}{x}$$

also holds, necessarily.

Since $x \in S$, we know $x > 0$ and so, certainly, $x + 2 > 0$, as well. Thus, we can multiply both sides of the known inequality by $x + 2$, yielding

$$x - 3 > (x+2)\left(1 - \frac{1}{x}\right) = x + 2 - \frac{x+2}{x}$$

Adding $3 + \frac{x+2}{x}$ to both sides, subtracting 2 from both sides, and rewriting in the reverse direction (for ease of reading), we obtain

$$x + 3 < x - 2 + \frac{x+2}{x}$$

Since $x - 2 < x + 2$, we deduce that

$$x + 3 < x + 2 + \frac{x+2}{x}$$

and factoring tells us

$$x + 3 < (x+2)\left(1 + \frac{1}{x}\right)$$

Again, since $x + 2 > 0$, we can divide both sides by $x + 2$, obtaining

$$\frac{x+3}{x+2} < 1 + \frac{1}{x}$$

which was the desired inequality. This shows $P(x) \implies Q(x)$, and since x was arbitrary, we have proven the claim. □

A key lesson here lies in how we took our scratch work and presented it in a different way in our proof. We cut out the unnecessary simplification and re-factoring steps, but we also noted why each step was valid as we performed it. A more seasoned mathematician would likely skip several of these steps and leave it to the reader to verify the algebraic work, but since we are early on in our mathematical careers, we thought it would be prudent to show as many details as possible.

Contrapositive Method

Look back to Section 4.6.1. There, we proved that a conditional statement is logically equivalent to its contrapositive. That is, the conditional statement

$$P \implies Q$$

necessarily has the same truth value as the statement

$$\neg Q \implies \neg P$$

For this reason, when we try to prove $P \implies Q$ is valid, we can just prove that $\neg Q \implies \neg P$ is valid, instead! Depending on what P and Q are, maybe this contrapositive is easier to understand, or we can spot a proof more quickly. In fact, the contrapositive strategy is particularly useful when P (or Q, or maybe both) has a "not" in it somewhere; by considering its negation, we can work instead with a "positive" assertion, instead of a negation.

Strategy:

Claim: $P \implies Q$

Contrapositive proof strategy:

Suppose that $\neg Q$ holds. Prove that $\neg P$ holds.

(Notice that this is the *direct proof strategy* applied to $\neg Q \implies \neg P$.)

Example 4.9.9. *Even products of integers:*

Statement: Let $E(x)$ be the proposition "x is even".

We claim that

$$\forall m, n \in \mathbb{Z}\,.\; E(m \cdot n) \implies \big(E(m) \vee E(n)\big)$$

Said another way, whenever the product of two integers is even, this necessarily means that at least one of the integers is even.

Implementation:

Proof. We prove this by contrapositive.

Let $m, n \in \mathbb{Z}$ be arbitrary and fixed.

Suppose that $\neg E(m) \wedge \neg E(n)$.

This means that m is odd and n is odd.

This means $\exists k, \ell \in \mathbb{Z}\,.\; m = 2k+1 \wedge n = 2\ell + 1$.

Let such k, ℓ be given. Then,

$$m \cdot n = (2k+1)(2\ell+1) = 4k\ell + 2k + 2\ell + 1 = 2(2k\ell + k + \ell) + 1$$

Since $2k\ell + k + \ell \in \mathbb{Z}$, as well, this shows that $m \cdot n$ is odd.

Thus, $\neg E(m \cdot n)$ holds, so we have shown that

$$\big(\neg E(m) \wedge \neg E(n)\big) \implies \neg E(m \cdot n)$$

The claim follows by contrapositive. □

Notice that we pointed out for our reader, at the beginning of the proof, that we would be using the contrapositive method. If we don't do that, the reader might be confused! "Why are we supposing that $\neg E(m)$ holds? What good is that?!", our reader might wonder. By revealing our strategy beforehand, we ensure that our reader will be able to follow along, avoiding unnecessary bewilderment.

Indirect Method (Proof By Contradiction)

This method depends on the logical negation of conditional statements. Reread Section 4.7 to see where we proved that

$$\neg(P \implies Q) \iff (P \wedge \neg Q)$$

This proof technique makes use of this equivalence.

Strategy:

Claim: $P \implies Q$

Indirect proof strategy:

> AFSOC that P holds and suppose that Q fails. Find a contradiction.

Example 4.9.10. *A surprising form of the AGM Inequality:*

Statement: $\forall x \in \mathbb{R}\,.\; x > 0 \implies x + \frac{1}{x} \geq 2$

Let's jump right into this proof without doing any scratch work, because we think this proof reads fairly straightforwardly. Afterwards, we'll discuss an alternate strategy.

Implementation:

Proof. Let $x \in \mathbb{R}$ be arbitrary and fixed.

Suppose $x > 0$.

AFSOC that $x + \frac{1}{x} < 2$.

Since $x > 0$, we can multiply through this inequality by x, yielding

$$x^2 + 1 < 2x$$

Subtracting and factoring, we obtain

$$(x-1)^2 < 0$$

This is a contradiction, since $(x-1)^2 \geq 0$. ⨳

Thus, our original assumption is invalid, and the claim follows. □

Now, you might be wondering about the title for this example. What does this have to do with the AGM Inequality? (Recall that we proved that fact back in Section 4.9.2.) An astute reader will possibly recognize that not only is this fact here an inequality (like the AGM), but also a couple of steps in this proof are similar to what we did to prove the AGM Inequality. Specifically, to prove the AGM Inequality, we started by using the fact that a particular squared expression is non-negative. Likewise, in this proof, we appealed to the fact that a squared expression *should* be non-negative. This similarity between the proofs indicates some potential underlying relationship. Indeed, we can actually directly *apply* the AGM Inequality (in a clever way, mind you!) to prove the above fact in a different way.

Think about this for a few minutes and see if you can come up with the following proof, before we show you how it works. What does it mean to *apply* the AGM Inequality? That result holds for any x and y, but here we have just one x. Can we be crafty about choosing what y should be so that the result here just "falls out" immediately? Try it! Then, read on ...

Proof. Let $x \in \mathbb{R}$ be arbitrary and fixed. Suppose $x > 0$.

Define $y = \frac{1}{x}$, so $y \in \mathbb{R}$.

Then, the AGM inequality applies to x and y (since that fact holds for *arbitrary* $x, y \in \mathbb{R}$). This tells us that

$$x \cdot \frac{1}{x} \leq \left(\frac{x + \frac{1}{x}}{2} \right)^2$$

Simplifying both sides slightly yields

$$1 \leq \frac{1}{4}\left(x + \frac{1}{x}\right)^2$$

and then multiplying both sides by 4 yields

$$4 \leq \left(x + \frac{1}{x}\right)^2$$

Since both sides are non-negative, we can take the square root of both sides and deduce that

$$2 \leq x + \frac{1}{x}$$

This proves the claim. □

There's a lesson here:

> Always be on the lookout for similarities between arguments and proofs, not just the results that are proven.

You might be able to save yourself some work by applying another result that has already been proven! (In this case, we didn't save ourselves too much writing; however, we might have saved ourselves some time, if we hadn't already noticed that the contradiction method would be fruitful. In particular, we might not have thought of the factoring trick that comes up in our first proof.)

4.9.6 Proving $\iff$ Claims

Recall that the "$\iff$" connective is defined entirely in terms of the "$\implies$" connective. That is, asserting that

$$P \iff Q$$

is logically equivalent to asserting two conditional statements:

$$(P \implies Q) \wedge (Q \implies P)$$

This gives rise to an obvious strategy: prove one conditional statement, then prove the other! The most common mistake we notice is when someone simply proves one statement or the other, but not both. Always keep that in mind!

Direct Method

Strategy:

Claim: $P \iff Q$

Direct proof strategy:

Prove that $P \implies Q$ (using one of the methods above).

Prove that $Q \implies P$ (using one of the methods above).

Example 4.9.11. *Even squares of integers:*

Statement: An integer is even if and only if its square is even.

Let's rewrite this claim using logical symbolic notation.

Let $E(z)$ be the proposition "z is even". Then we claim that

$$\forall z \in \mathbb{Z}. \left(E(z) \iff E(z^2)\right)$$

Implementation:

Proof. Let $z \in \mathbb{Z}$ be arbitrary and fixed.

($\implies$) First, assume z is even, so $\exists k \in \mathbb{Z}.\ z = 2k$. Let such a k be given.

Since $z = 2k$, we can square both sides and obtain

$$z^2 = (2k)^2 = 4k^2 = 2(2k^2)$$

Define $\ell = 2k^2$. Notice that $\ell \in \mathbb{Z}$ and $z^2 = 2\ell$

This shows that z^2 is even.

Thus, $E(z) \implies E(z^2)$.

($\impliedby$) Second, we will prove $E(z^2) \implies E(z)$ by contrapositive.

Suppose z is odd, so $\exists m \in \mathbb{Z}.\ z = 2m + 1$. Let such an m be given.

Since $z = 2m + 1$, we can square both sides and obtain

$$z^2 = (2m + 1)^2 = 4m^2 + 4m + 1 = 2(2m^2 + 2m) + 1$$

Define $n = 2m^2 + 2m$. Notice that $n \in \mathbb{Z}$ and $z^2 = 2n + 1$.

This shows that z^2 is odd.

Thus, $\neg E(z) \implies \neg E(z^2)$; by contrapositive, then, $E(z^2) \implies E(z)$.

Since we have shown both directions, we conclude that

$$E(z) \iff E(z^2)$$

and since z was arbitrary, this biconditional holds for all integers z. $\square$

Indirect Method (Proof by Contradiction)

Strategy:

Claim: $P \iff Q$

Indirect proof strategy:

AFSOC that $\neg(P \implies Q) \vee \neg(Q \implies P)$.

Consider the first case, where $P \wedge \neg Q$ holds. Find a contradiction.

Consider the second case, where $Q \wedge \neg P$ holds. Find a contradiction.

Implementing this strategy—and deciding when to do so, even—depends on the actual statements P and Q. In general, a direct method will probably be better (not always), but if you find yourself getting stuck, consider looking at the negations—$P \wedge \neg Q$, and $Q \wedge \neg P$—and see if that takes you anywhere. It's worth a try!

Intermediary Method (TFAE)

For lack of a better term, we are going to call this strategy an **intermediary method**. As you'll see, it's not exactly a direct method, but neither is it an indirect method. In implementing this strategy, we don't have to look at any logical negations, but we also aren't directly linking the statements P and Q.

Rather, this method requires us to find some *intermediary* statement R and proving two biconditional statements: namely, $P \iff R$ and $R \iff Q$. This yields the following chain of conditional statements

$$P \iff Q \iff R$$

which tells us that all three statements have the same truth value. In particular, then, P and Q must always have the same truth value, so we conclude that $P \iff Q$.

The acronym **TFAE** stands for "the following are equivalent". We chose this name because it is a common phrase in mathematics; it is used in theorems that present a list of conditions/properties that all "imply each other". That

is, some theorems list several properties and assert that all of them are logically equivalent, whence "the following are equivalent". To prove such a theorem, one would implement the above strategies over and over and prove that the statements are, indeed, equivalent. The only difference here is that we have to *come up with* the intermediary statement to use. (But hey, whoever presented and proved a TFAE-style theorem probably had to come up with all of those statements to begin with, too!)

Strategy:

Claim: $P \iff Q$

Intermediary strategy:

> Define a statement R.
>
> Prove that $P \iff R$ (using one of the methods above). Prove that $Q \iff R$ (using one of the methods above).

4.9.7 Disproving Claims

We have now discussed (and seen, in many examples) how to **prove** any kind of mathematical statement. Fantastic! But you might be saying, "Uh oh . . . What if I want to **disprove** a statement?" Our answer to that question is short and sweet: *there's no difference.*

To **disprove** a statement means you want to show its truth value is False. By the definition of logical negation, this means you want to show that the statement's negation has the truth value True. Accordingly, you can just find and write down that logical negation and prove that statement to be True, using any of the strategies we have explored in this section. Voilá!

Just for the sake of illustration let's see an example of this phenomenon in action. Specifically, we'll see an example where we want to disprove a "$\forall$" claim, meaning we want to prove a "$\exists$" claim. This is where the notion of a **counterexample** comes into play.

Counterexamples

In general, a **counterexample** is an instance of a statement that disproves a universal quantification. It's an *example* because it works to prove an "$\exists$" claim, and it's *counter* in the sense that it shows this specific examples does *not* have the claimed property.

Example 4.9.12. Look back to Example 4.9.8. In it, we defined the set

$$S = \{x \in \mathbb{R} \mid x > 0\}$$

and then defined two variable propositions:

$$P(x) \text{ is } \text{“ } \frac{x-3}{x+2} > 1 - \frac{1}{x} \text{ ”}$$
$$Q(x) \text{ is } \text{“ } \frac{x+3}{x+2} < 1 + \frac{1}{x} \text{ ”}$$

Then, we proved that

$$\forall x \in S.\ P(x) \implies Q(x)$$

In this example, we will consider the statement

$$\forall x \in S.\ Q(x) \implies P(x)$$

Specifically, we will **disprove** it. Before we do that, though, play around with the statement on your own. Try to prove it, even though we essentially just told you it's False! Do you find your "proof" breaking down somewhere? Why is that happening? Can you use your observation to help you find a counterexample to the claim? See what you can find, then read on.

Scratch work:

To start, we need the logical negation of the claim we are disproving:

$$\exists x \in S.\ Q(x) \wedge \neg P(x)$$

This means we need to find a specific real number x that satisfies three conditions: (1) the inequality $x > 0$, (2) the inequality

$$\frac{x+3}{x+2} < 1 + \frac{1}{x}$$

and (3) the inequality

$$\frac{x-3}{x+2} \leq 1 - \frac{1}{x}$$

There are a few strategies we could use. Like we mentioned above, we could try to (erroneously, of course) prove that the first inequality does, indeed, imply the second one, and ascertain where this breaks down. Alternatively, we could "try some values" using an "educated guess" method.

Knowing that $x \in \mathbb{R}$ and $x > 0$ indicates, to us anyway, that we might want to try "extreme" values of x. This means either "small" x (i.e. x close to 0) or "large" x (i.e. ever-increasing values of x, that grow larger until we find one that works).

It seems easier to first work with some "small" values, so let's try $x = 1$. We see that (1) holds because $1 > 0$, and (2) holds because $\frac{4}{3} < 2$, and (3) holds because $-\frac{2}{3} < 0 \leq 0$. Cool, that's it!

Proof. Here, we will disprove the claim that $\forall x \in \mathbb{R}\,.\; Q(x) \implies P(x)$.

Consider $x = 1$. Notice that $x \in \mathbb{R}$ and $x > 0$.

Also, notice that $Q(1)$ holds because

$$\frac{1+3}{1+2} = \frac{4}{3} < 2 = 1 + \frac{1}{1}$$

Also, notice that $P(1)$ fails because

$$\frac{1-3}{1+2} = -\frac{2}{3} \not> 0 = 1 - \frac{1}{1}$$

Thus, we have shown that

$$\exists x \in S\,.\; Q(x) \wedge \neg P(x)$$

and this disproves the claim. □

4.9.8 Using assumptions in proofs

Another important aspect of creating and writing formal proofs is that we are sometimes given **assumptions** to use. When we state a theorem, it usually has some **hypotheses** and a **conclusion**. We get to temporarily add those assumptions to our mathematical toolkit; we can use them to get to the desired conclusion. Similarly, along the way, we might develop some further facts and observations, and we can keep those with us and use them to prove the conclusion, as well. In this short section, we want to point out three observations and issues that may come up while you are *using* an assumption in a proof.

"$P \vee Q$" means Use Cases

If, at some point in a proof, you have assumed or deduced that $P \vee Q$ holds, how can you proceed? Knowing this disjunction holds means that at least one of the constituent statements—P or Q—holds. Thus, you can consider each of those two **cases** separately. For example, your proof might have this section in it:

> Because $P \vee Q$, We have two cases.
>
> > **Case 1:** Suppose P holds. Then ...
> >
> > **Case 2:** Suppose Q holds. Then ...

As long as you can achieve your desired goal in both cases, you can make that deduction.

Notice that there is no need to consider the case where *both* P and Q hold. For one, this might not necessarily even happen. But also, if you only end up using one or the other of the statements to deduce your desired conclusion, then there was no need to temporarily assume both of them.

We have been using cases all along in some of our proofs. Now, we see exactly why they work! We use cases when there is an underlying disjunction of statements.

"There exists ..." vs. "Let ... be given"

This is a subtle but important distinction. If you write down a claim like

$$\exists x \in S \text{ . } P(x)$$

in the middle of a proof, what have you asserted? Technically speaking, you have really only stated that the line above is a True statement; you have asserted that there *does* exist some $x \in S$ with the property $P(x)$. But, if you move and start referring to x afterwards ... this is not valid! Nowhere in the assertion of *existence* did you introduce a *particular instance* of that claim. It might be the case that several such x elements exist. Do you want to talk about all of them? Or just a particular one? Don't leave it up to the reader of your proof to intuit exactly what you're doing!

If you know, or have assumed, some existence statement (like the line above) and you want to actually *introduce* a variable that satisfies that existence claim, use the following wonderful phrase:

> "Let such an x be given."

This signals to the reader that not only are you saying such an x exists, but you are also bringing it into play in your proof. You want the letter x, for the rest of your written argument, to represent an element with that property. Thereafter, you get to refer to that object x by name.

If you assert the existence of several variables and want to introduce them, just use a similar phrase with a slightly different verb. For instance, you might write something like this:

> ... and so we deduce that $\exists x, y, z \in \mathbb{Z}$ such that $P(x, y, z)$ holds.
>
> Let such x, y, z be given. Observe that ...

"$P \implies Q$" vs. "P, therefore Q"

This distinction hinges on an idea similar to the previous example we just mentioned. Specifically, there is a difference between writing a statement to assert its *validity* and writing a statement to show the reader you are making a *conclusion* from it. In the last example, this was the distinction between saying something exists versus introducing such an object.

Here, the distinction lies between asserting a conditional statement—like $P \implies Q$—to say that this conditional relationship exists versus using this statement to *deduce* that Q holds. Technically speaking, just writing "$P \implies Q$" on your paper does not assert that Q is valid. You must make it very clear to your reader that you *also* know P and are using the conditional statement to *deduce* Q.

Look back at our discussion in Section 4.5.6. There, we described this important distinction, and referred to the method of "Modus ponens". As we mentioned, if you want to actually deduce Q, you should write something like:

> $P \implies Q$ because ...
>
> Also, P holds because ...
>
> Therefore, Q holds.

4.9.9 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the Direct Method for an $\exists$ claim? What are the important steps in showing an object exists?

(2) What is the Direct Method for an $\implies$ claim? How do we prove a $\implies$ claim by contradiction? How are these methods different?

(3) How do we prove a $\iff$ claim?

(4) What is the AGM Inequality? Where does the acronym come from?

(5) To which type of claim does the Contrapositive Strategy apply? Why does it work?

(6) What is a counterexample?

(7) What is the difference between saying "$\exists a \in A.\ P(a)$" and saying "$\exists a \in A.\ P(a)$ so let such an a be given"?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Prove that $\forall x \in \mathbb{R}.\ x^2 \neq 1 \implies x \neq 1$.

(2) Prove that $\forall n \in \mathbb{N}.\ n \geq 5 \implies 2n^2 > (n+1)^2$

(3) Express the following claim in logical symbols, and then prove it.

There exists an even natural number that can be written as the sum of two primes in two *different* ways.

(4) Prove that every natural number is either less than $\sqrt{10}$ or bigger than 3. That is, prove

$$\forall n \in \mathbb{N}\,.\; n < \sqrt{10} \vee n > 3$$

(5) Let A, B, C, D be sets. Prove that, if $A \cup B \subseteq C \cup D$ and $C \subseteq A$ and $A \cap B = \varnothing$, then $B \subseteq D$.

(6) Define $P = \{y \in \mathbb{R} \mid y > 0\}$. Prove that

$$\forall \varepsilon \in P\,.\; \forall x, y \in \mathbb{R}\,.\; \exists \delta \in P\,.\; |x - y| < \delta \implies |(3x - 4) - (3y - 4)| < \varepsilon$$

Can you also prove the following claim? How is it different than the one above?

$$\forall \varepsilon \in P\,.\; \exists \delta \in P\,.\; \forall x, y \in \mathbb{R}\,.\; |x - y| < \delta \implies |(3x - 4) - (3y - 4)| < \varepsilon$$

(7) Let $E(x)$ be the proposition "x is even". Prove that

$$\forall a, b \in \mathbb{Z}\,.\; E(a) \wedge E(b) \iff E(a + b) \wedge E(a \cdot b)$$

(8) Look back at the proof that $\sqrt{2}$ is irrational. Modify it to provide a proof that $\sqrt{3}$ is irrational, as well.

Try to do the same and prove that $\sqrt{6}$ is irrational.

Challenge: Can you prove that $\sqrt{p}$ is irrational for *every* prime number p?

(9) Prove that there are infinitely many rational numbers.

Hint: Do this by contradiction. (Suppose there are finitely many ...)

4.10 Summary

We now have a huge toolkit of mathematical proof strategies! We developed the necessary terminology and symbols to properly express mathematical claims. Then, we used the concepts underlying those to develop proof strategies and saw many examples of how they are used.

One of the more difficult aspects of *writing* a proof is figuring out the proof in the first place! This involves not only figuring out whether a claim is True or False, but also ultimately deciding which proof strategy to implement. We realize this is challenging because, well, it is ... Furthermore, it's difficult to characterize exactly *when* to use certain strategies. We can offer guidelines and suggestions (and have done so as much as possible) but, in the long run, the best way to get *better* at implementing proof strategies and deciding which ones to use is to *just do it*. Approach an exercise problem. Try to play with the statement and see why it is True (or False, as the case may be). Try using a

proof strategy on it. Did it work? How? If it didn't, why and where did it break down? Can you use those observations to decide on a different approach to the problem? Ultimately, can you write down a good, formal argument? Working through these steps on as many problems as possible truly is the best practice. There is no substitute for simply **doing mathematics**.

4.11 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them. Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ⋆.

Problem 4.11.1. ► Consider the universal context to be $U = \mathbb{Z}$.

Let $P(x)$ be the proposition "$1 \leq x \leq 3$".

Let $Q(x)$ be the proposition "$\exists k \in \mathbb{Z}\text{.}\ x = 2k$".

Let $R(x)$ be the proposition "$x^2 = 4$".

Let $S(x)$ be the proposition "$x = 1$".

For each of the following statements, write out an English sentence to describe what the statement means, then write the logical negation, and then decide which claim is True or False, and why.

(a) $\forall x \in \mathbb{Z}\text{.}\ P(x) \implies Q(x)$

(b) $\exists x \in \mathbb{Z}\text{.}\ R(x) \wedge P(x)$

(c) $\forall x \in \mathbb{Z}\text{.}\ R(x) \implies P(x)$

(d) $\forall x \in \mathbb{Z}\text{.}\ \exists y \in \mathbb{Z}\text{.}\ x \neq y \wedge P(x) \wedge R(y)$

(e) $\forall x \in \mathbb{Z}\text{.}\ \exists y \in \mathbb{Z}\text{.}\ (S(x) \vee Q(x)) \wedge P(y) \wedge \neg Q(y)$

(f) $\exists x \in \mathbb{Z}\text{.}\ S(x) \iff P(x) \wedge \neg Q(x)$

(g) $\exists x \in \mathbb{Z}\text{.}\ S(x) \iff \neg P(x) \wedge Q(x)$

Problem 4.11.2. For each of the following claims, define some sets and variable propositions to express the claim in concise symbolic, logical notation. Then, write the logical negation, as well. Note which one is True or False.

(a) Every odd natural number is prime.

(b) There is a real number that is strictly greater than any integer squared.

(c) Some real number between -1 and 1 has the property that it is equal to the cube of some *different* real number between -1 and 1.

(d) The union of the sets of multiples of primes is the set of natural numbers itself.

Problem 4.11.3. Consider the following defined sets and questions about those sets. For each question, if your answer is No, provide an example to demonstrate this.

(a) Let $S = \{1, 2, 3, 4\}$ and $T = \{3, 4, 5, 6, 7, 8\}$.

Is it true that $\forall s \in S\,.\ \exists t \in T\,.\ s + t = 7$?

(b) Let $S = \{2, 3, 4, 5, 6\}$ and $T = \{3, 4, 5, 6\}$.

Is it true that $\forall s \in S\,.\ \exists t \in T\,.\ s + t = 7$?

(c) Let $S = \mathbb{N}$ and $T = \mathbb{Z}$.

Is it true that $\forall s \in S\,.\ \exists t \in T\,.\ s + t = 7$?

(d) Let $S = \mathbb{Z}$ and $T = \mathbb{N}$.

Is it true that $\forall s \in S\,.\ \exists t \in T\,.\ \ s + t = 7$?

Problem 4.11.4. Consider the following defined sets and questions about those sets. For each question, if your answer is No, provide an example to demonstrate this.

(a) Let $S = \{1, 2, 3\}$, $T = \{6, 7, 8, 9\}$, and $V = \{7, 8, 9, 10\}$.

Is it true that $\exists s \in S\,.\ \forall t \in T\,.\ \exists v \in V\,.\ s + t = v$?

(b) Let $S = \{1, 2, 3\}$, $T = \{4, 5, 6, 7\}$, and $V = \{5, 6, 7, 9, 10, 11\}$.

Is it true that $\exists s \in S\,.\ \forall t \in T\,.\ \exists v \in V\,.\ s + t = v$?

(c) Let $S = T = V = \mathbb{N}$.

Is it true that $\exists s \in S\,.\ \forall t \in T\,.\ \exists v \in V\,.\ s + t = v$?

(d) Let $S = \mathbb{N}$, $T = \mathbb{Z}$, and $V = \mathbb{N}$.

Is it true that $\exists s \in S\,.\ \forall t \in T\,.\ \exists v \in V\,.\ s + t = v$?

Problem 4.11.5. Prove or disprove the following claim rigorously:

$$\exists x \in \mathbb{R}\,.\ \forall y \in \mathbb{R}\,.\ x^2 - y^2 \geq 0$$

Problem 4.11.6. Prove or disprove the following claim rigorously:

$$\forall x, y \in \mathbb{Z}\,.\ \exists z \in \mathbb{N} \cup \{0\}\,.\ ((x - y) = z) \vee ((y - x) = z)$$

Problem 4.11.7. Prove that there are no integral solutions (i.e. $x, y \in \mathbb{Z}$) to the equation $x^2 - y^2 = 14$.

Problem 4.11.8.

Problem 4.11.9.

Problem 4.11.10.

Problem 4.11.11.

Problem 4.11.12. Use *logical equivalences* to prove that

(a) $(A \cup B) \cap \overline{A} = B - A$

(b) $A \cap (B - C) = (A \cap B) - (A \cap C)$

Problem 4.11.13. Define the sets

$$A = \left\{ (x, y) \in \mathbb{R} \times \mathbb{R} \mid \frac{x}{y} + \frac{y}{x} \geq 2 \right\}$$

and

$$B = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid (x > 0 \wedge y > 0) \vee (x < 0 \wedge y < 0)\}$$

Is it the case that $A \subseteq B$? If so, prove it. Otherwise, exhibit a counterexample.

Is it the case that $B \subseteq A$? If so, prove it. Otherwise, exhibit a counterexample.

Problem 4.11.14. Let $P = \{y \in \mathbb{R} \mid y > 0\}$ be the set of positive real numbers. Prove the following claim:

$$\forall \varepsilon \in P\,.\ \exists \delta \in P\,.\ \forall x \in \mathbb{R}\,.\ |x| < \delta \implies |x^2| < \varepsilon$$

Problem 4.11.15. What is wrong with the folllowing "proof" of the claim that $\mathcal{P}(C \cup D) = \mathcal{P}(C) \cup \mathcal{P}(D)$?

> Let $X \in \mathcal{P}(C \cup D)$ be arbitrary and fixed.
>
> This means $X \subseteq C \cup D$.
>
> So, $X \subseteq C \vee X \subseteq D$.
>
> Then, $X \in \mathcal{P}(C) \vee X \in \mathcal{P}(D)$.
>
> Thus, $X \in \mathcal{P}(C) \cup \mathcal{P}(D)$.
>
> Therefore, $\mathcal{P}(C \cup D) = \mathcal{P}(C) \cup \mathcal{P}(D)$.

Problem 4.11.16. Suppose $x \in \mathbb{Z}$ and x^2 is a multiple of 8. Prove that x is even.

Is the converse of this statement True? If so, prove it; otherwise, exhibit a counterexample.

Problem 4.11.17. Define the proposition $E(z)$ to be "z is even. Prove that

$$\forall z \in \mathbb{Z}\text{.}\ E(z) \iff E(z^3)$$

Problem 4.11.18. Use the result of Problem 4.11.17 to prove that $\sqrt[3]{2}$ is irrational.

Problem 4.11.19. Let $P = \{y \in \mathbb{R} \mid y > 0\}$. Prove that

$$\bigcap_{x \in P} \left\{ y \in \mathbb{R} \mid 1 - \frac{1}{x} < y < 2 \right\} = \{z \in \mathbb{R} \mid 1 < z < 2\}$$

Problem 4.11.20. Let $A, B, C \subseteq U$ be sets. Define

$$S = \big((A \cap \overline{B}) \cup C\big) - A$$

and

$$T = C - (A \cup B)$$

Is $S \subseteq T$? If so, prove it; otherwise, find a counterexample.

Is $T \subseteq S$? If so, prove it; otherwise, find a counterexample.

Problem 4.11.21. For each $x \in \mathbb{R}$, define the set

$$S_x = \{y \in \mathbb{R} \mid -x \leq y \leq x\}$$

Also, define the set

$$P = \{y \in \mathbb{R} \mid y > 0\}$$

Prove each of the following claims.

$$\bigcap_{x \in P} S_x = \{0\}$$

$$\bigcap_{x \in \mathbb{N}} S_x = \{y \in \mathbb{R} \mid -1 \leq y \leq 1\}$$

Problem 4.11.22. Let $P(x)$ be the variable proposition

$$\text{“ } \frac{x^2 + 4}{x^2 + 1} < 1 + \frac{1}{x} \text{ ”}$$

and let $Q(x)$ be the variable proposition

$$\text{“ } \frac{x^2 - 4}{x^2 + 1} > 1 - \frac{1}{x} \text{ ”}$$

Also, let $S = \{x \in \mathbb{R} \mid x > 0\}$ be the set of positive real numbers.

For each of the following statements determine whether it isTrue (in which case, provide a proof) or False (in which case, provide a counterexample and demonstrate why it is valid).

(a) $\forall x \in S\,.\; P(x) \implies Q(x)$

(b) $\forall x \in S\,.\; Q(x) \implies P(x)$

Problem 4.11.23. Let A and B be any two sets. Prove that

$$A \times B = B \times A \iff (A = B \;\vee\; A = \varnothing \;\vee\; B = \varnothing)$$

(Don't forget that this is an *if and only if* claim!)

Problem 4.11.24. Let A, B, C, D be any sets. Prove that

$$(A \times B) \cap (C \times D) = \varnothing \iff \big(A \times B = \varnothing \vee C \times D = \varnothing\big)$$

Problem 4.11.25. Let B be any set. Let I be an index set, and let A_i be a set for every $i \in I$. Prove the following set equalities:

(a) $\left(\bigcap_{i\in I} A_i\right) - B = \bigcap_{i\in I}(A_i - B)$

(b) $\left(\bigcup_{i\in I} A_i\right) - B = \bigcup_{i\in I}(A_i - B)$

(c) $\left(\bigcap_{i\in I} A_i\right) \times B = \bigcap_{i\in I}(A_i \times B)$

(d) $\left(\bigcup_{i\in I} A_i\right) \times B = \bigcup_{i\in I}(A_i \times B)$

(e) $B - \left(\bigcap_{i\in I} A_i\right) = \bigcup_{i\in I}(B - A_i)$

(f) $B - \left(\bigcup_{i\in I} A_i\right) = \bigcap_{i\in I}(B - A_i)$

Problem 4.11.26. In this problem, you will prove that the rational numbers $\mathbb{Q}$ are **dense**. Namely, we want you to consider the following proposition:

Proposition. Strictly between any two distinct rational numbers lies another rational number.

Restate this claim using logical symbols, and then **prove** it.

Problem 4.11.27. This problem is meant to introduce the concept of **uniqueness**. We say an object with a certain property is **unique** if it has the desired property but no **other** object has that property.

That is, we would say x is the unique element of S with property $P(x)$ if and only if

$$\exists x \in S \,.\, P(x) \wedge (\forall y \in S \,.\, y \neq x \implies \neg P(y))$$

Notice that this is logically equivalent to

$$\exists x \in S \,.\, P(x) \wedge (\forall y \in S - \{x\} \,.\, \neg P(y))$$

Also, we can write the contrapositive instead:

$$\exists x \in S \,.\, P(x) \wedge (\forall y \in S \,.\, P(y) \implies x = y)$$

Use this to restate the following claim in logical symbols. Then, **prove** it.

Claim: There is a unique *natural* root of the equation $n^3 - n - 6 = 0$.

Problem 4.11.28. This problem provides a definition of a new set operation (defined in terms of others) and asks you to prove several set containments and equalities, using this operation.

Definition: Let A, B be sets. The **symmetric difference** of A and B is denoted by $A \Delta B$ and is defined as

$$A \Delta B = (A - B) \cup (B - A)$$

Now, let A, B, C be any sets. Prove the following:

(a) $A \Delta A = \varnothing$

(b) $A \Delta B = B \Delta A$

(c) $A \Delta \varnothing = \varnothing$

(d) $A \subseteq B \implies A \Delta B = B - A$

(e) $A \Delta (B \Delta C) = (A \Delta B) \Delta C$

(f) $\overline{A} \Delta \overline{B} = A \Delta B$ (supposing $A, B \subseteq U$)

(g) $(A \Delta B) \cap C = (A \cap C) \Delta (B \cap C) = (A \Delta B) - C$

Problem 4.11.29. In this problem, you will prove a useful result about **primality testing**. Much of modern cryptography is based on factoring large composite numbers into its prime factors. The following proposition says that we only need to check *up until* $\sqrt{p}$ for factors of p.

Proposition. Let p be a natural number that is at least 2. If none of the natural numbers between 2 and $\sqrt{p}$ (inclusive) divide p, then p is prime.

Recall: The formal definition of "|" is

> Given $a, b \in \mathbb{Z}$, we write $a \mid b$ if and only if $\exists k \in \mathbb{Z}\text{.}\ b = ak$.

Restate the above proposition using logical symbols. Then, **prove** it.

(**Hint:** Think about the contrapositive of the cliam ...)

Problem 4.11.30. Let A, B be any sets. Prove the following claim:

$$A \times B = B \times A \iff (A = B \vee A = \varnothing \vee B = \varnothing)$$

Problem 4.11.31. Let S, T be sets whose elements are also sets. For each of the following statements, either **prove** the statement holds in general, or provide a counterexample:

(a) $\displaystyle\bigcup_{X \in S \cup T} X \subseteq \left(\bigcup_{Y \in S} Y\right) \cup \left(\bigcup_{Z \in T} Z\right)$

(b) $\displaystyle\bigcup_{X \in S \cup T} X \supseteq \left(\bigcup_{Y \in S} Y\right) \cup \left(\bigcup_{Z \in T} Z\right)$

(c) $\displaystyle\bigcap_{X \in S \cup T} X \subseteq \left(\bigcap_{Y \in S} Y\right) \cap \left(\bigcap_{Z \in T} Z\right)$

(d) $\displaystyle\bigcap_{X \in S \cup T} X \supseteq \left(\bigcap_{Y \in S} Y\right) \cap \left(\bigcap_{Z \in T} Z\right)$

4.12 Lookahead

Now that we have all of these mathematical tools—and we've put them to good use, honing our skills with lots of exercsises—we are more equipped to step out into the mathematical wilderness. We will be exploring various branches of mathematics, learning some fundamental concepts and notation, and applying our proof strategies to new and wonderful results.

Before we do that, though, we need to take care of one lingering issue: we want to *formalize* induction. Before, we "waved our hands" a bit about what exactly induction is and how it works. We gave you the "Domino Analogy" and used it on some example. But now we have the requisite terminology and knowledge to properly describe mathematical induction and study its various forms. It might be a good idea to flip through Chapter 2 again to remind yourself of some of the illustrative examples where we used an inductive argument. Do you remember the Domino Analogy? Can you anticipate how we might formalize the Principle of Mathematical Induction? Can you explain that theorem to a friend and convince them? Try it! Then, read on!

Chapter 5

Rigorous Mathematical Induction: A Formal Restatement

5.1 Introduction

It might seem like we're being redundant by including this chapter after having already discussed mathematical induction. Our goals are many, though, and you will see afterwards why we have turned our eye backwards a bit to discuss this material once more.

First, we feel a little uncomfortable with how *informal* (mathematically speaking) we were with our initial treatment of induction. Second, we left a few lingering questions back in Chapter 2. What was different about some of the later examples we saw, like the Takeaway game and the Tower of Hanoi? Didn't they seem to use "more assumptions" in their inductive arguments than the other examples, like our proof that $\sum_{k\in[n]} k = \frac{n(n+1)}{2}$? We think so, and we will adress those differences here. Third, there are plenty of examples left to be seen that are interesting and useful facts in their own right, and working through them will help to develop our understanding of the mathematical language we are beginning to speak with each other. Fourth, the final theorem stated and proved (with your help!) in this chapter will be a striking example of *equivalence*; specifically, we will show that three theorems are all connected by biconditional statements! (This will be the first great example of a "The following are equivalent..."-style theorem, like we pointed out in the proof strategy for biconditional statements, in Section 4.9.6.)

5.1.1 Objectives

Since we have discussed induction before and are just returning to this topic, we will forego the usual introductory matter at the beginning of this chapter.

Instead, we will summarize the main objectives of this chapter here via a series of statements.

By the end of this chapter, you should be able to ...

- State the Principle of Mathematical Induction and describe how its proof is related to the set of natural numbers.
- State the Principle of Strong Mathematical Induction, compare and contrast this with the previous principle, and describe how it can be proven.
- Use inductive arguments to prove claims and, in particular, identify when a strong inductive argument is required.
- Understand and explain some variants of mathematical induction, and identify problems where these variants might be useful.
- State the Well-Ordering Principle and explain its relationship to mathematical induction.

5.2 Regular Induction

This first section concerns the kind of inductive arguments we have seen before. You will see why, in the next section, we would choose to refer to this as "Regular" Induction.

5.2.1 Theorem Statement and Proof

Here, let's recall the statement of the **Principle of Mathematical Induction** that we gave in Chapter 3. Think about how it follows the **Domino Analogy**, or whichever analogy works best to help you understand an inductive process. You might have missed this Theorem statement if you didn't complete the Optional Reading about Defining $\mathbb{N}$, in Section 3.8, but that's okay. We're confident you can still read this and formulate it in a way that corresponds to an inductive process.

Theorem 5.2.1 (Principle of Mathematical Induction)**.** *Let $P(n)$ be some "fact" or "observation" that depends on the natural number n. Assume that*

1. *$P(1)$ is a true statement.*
2. *Given any $k \in \mathbb{N}$, if $P(k)$ is true, then we can conclude* necessarily *that $P(k+1)$ is true.*

Then the statement $P(n)$ must be true for every *natural number $n \in \mathbb{N}$.*

Look at all of these wordy sentences and phrases and hand-wavey terms. Some "fact" that depends on a natural number? Sounds like a **variable proposition**, right? "If ... then we can conclude necessarily ..." Sounds like a **conditional statement**, doesn't it? All of this language is meant to express some logical underpinnings, and we can restate the whole theorem now, using the concepts and notation developed in the previous chapter. Try your hand at doing this first, before looking at our version. While you're at it, try remembering how we *proved* that theorem. (Again, you might have missed this if you skipped this optional reading, and that's fine.) Look back to Section 3.8.2 and remind yourself, because we will follow that same proof here, but we'll use the logical symbols and tools that we now have in hand. Ready? Here we go!

Theorem 5.2.2 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a variable proposition. Suppose that*

(1) $P(1)$ holds True*, and*

(2) $\forall k \in \mathbb{N}\,.\, P(k) \implies P(k+1)$ holds True

Then $\forall n \in \mathbb{N}\,.\, P(n)$ holds True*.*

That's all there is to it! This encapsulates all the same ideas—that some initial fact holds, and that every fact implies the next one, making all the facts hold—but it does so using logical symbols and language. Do you see how they say the same thing? Make sure you do before reading on!

Our goal now is to *prove* this theorem. Yes, we will prove that mathematical induction is a valid proof technique! Why shouldn't we? We proved (via a truth table) that a conditional statement is logically equivalent to its contrapositive, which gave us a proof strategy. Why shouldn't we prove this one, as well?

Before we show the proof, though, we want you to read the section on defining the natural numbers, Section 3.8. It contains the following key definitions, which we will make use of in the forthcoming proof. In that section, we defined what it means to be an **inductive set** and then stated that $\mathbb{N}$ is the "smallest" inductive set, in the sense that $\mathbb{N}$ is a *subset* of all the inductive sets in the universe. This is the property we wanted $\mathbb{N}$ to have, and these definitions made it so. We will give you those important definitions right here—rewritten slightly, using logical notation and foregoing some set theoretic concepts—but we also suggest that you read that section to grasp the full extent of the discussion.

Definition 5.2.3. *Let I be a set. If the following conditions hold:*

1. *$1 \in I$, and*
2. *For any element k, the implication $k \in I \implies k+1 \in I$ holds;*

then I is called an **inductive set***.*

Definition 5.2.4. *The set of all* ***natural numbers*** *is the set*

$$\mathbb{N} := \{x \mid \text{ for every inductive set } I,\ x \in I\}$$

Put another way, $\mathbb{N}$ *is the* smallest *inductive set:*

$$\mathbb{N} = \bigcap_{I \in \{S | S \text{ is inductive}\}} I$$

Okay, now we're ready for the proof!

Proof. Let $P(n)$ be a variable proposition, defined for every natural number n. Suppose that the two conditions given in the theorem do hold, namely

(1) $P(1)$ holds True, and

(2) $\forall k \in \mathbb{N}.\ P(k) \implies P(k+1)$ holds True

Let S be the set of instances for which $P(n)$ is True. That is, define

$$S = \{n \in \mathbb{N} \mid P(n) \text{ is } \mathsf{True}\}$$

By definition (using set-builder notation), $S \subseteq \mathbb{N}$.

Condition (1) above guarantees that $1 \in S$.

Condition (2) above guarantees that $\forall k \in \mathbb{N}.\ k \in S \implies k+1 \in S$.

Together, these two conditions guarantee that S is an *inductive* set. By the definition of $\mathbb{N}$ above, we therefore know that $\mathbb{N} \subseteq S$.

Thus, by a double-containent argument $S = \mathbb{N}$. This means that the statement $P(n)$ holds for *every* natural number n, i.e. $\forall n \in \mathbb{N}.\ P(n)$ is True! □

Understanding the set-theoretic mechanics behin this proof are not essential to being able to use induction and write inductive proofs. However, we believe that thinking about these logical underpinnings can only help your understanding, or spark some curiosity in mathematical logic and set theory, or possibly both!

The important thing that we have accomplished here, by restating the PMI, is that we now have a clear way of determining whether an inductive argument has succeeded. The entire crux of a "proof by induction" lies in verifying conditions (1) and (2) in the statement of the theorem (i.e. verifying that the "truth set" for the proposition $P(n)$ is an inductive set).

5.2.2 Using Induction: Proof Template

Taking the observation above, we can develop a proof template for a proper **"proof by induction"**. (This can be added to the list of proof strategies from the last chapter, as well, thereby broadening our mathematical toolkit!) Notice that all of the steps in this template are motivated by making our proof readable, orderly, and logically correct:

- We must define a proposition $P(n)$to show our reader what we aim to prove.

- We must verify the **Base Case (BC)** to show that condition (1) in the PMI is satisfied.
- We must verify the the conditional statement $\forall k \in \mathbb{N}.\ P(k) \implies P(k+1)$ to show that condition (2) in the PMI is satisfied. To do this, we will apply the direct proof strategy for proving conditional statements; this has two parts:
  - First, we make an **Inductive Hypothesis (IH)**, which introduces an arbitrary and fixed natural number k and supposes $P(k)$ holds.
  - Second, we go through the **Inductive Step (IS)**, which takes that assumption and deduces that $P(k+1)$, also holds.
- Between these steps—the **BC**, the **IH**, and the **IS**—we have verified the conditions of PMI, and we can deduce its conclusion: $\forall n \in \mathbb{N}.\ P(n)$.

  Finally, we make this conclusion to remind our reader of what we have accomplished.

Template for a "Proof by Induction"

Goal: Prove that $\forall n \in \mathbb{N}.\ P(n)$

Proof.

Let $P(n)$ be the proposition " _____ ".

We will prove $\forall n \in \mathbb{N}.\ P(n)$ by induction on n.

Base Case: Observe that $P(1)$ holds because _____ .

Induction Hypothesis: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $P(k)$ holds.

Induction Step: Deduce that $P(k+1)$ also holds.

By PMI, it follows that $\forall n \in \mathbb{N}.\ P(n)$. □

Comments and Common Pitfalls

What follows are some recommendations and suggestions. These are based on what we feel constitutes a good, well-written inductive argument, and also some mistakes that we see students consistently make over the years.

- **Be sure to *define* a proposition.**

  Sometimes the claim is defined for you in the statement of a problem or exercise. However, it is not always defined explicitly as $P(n)$. In that case, referring to a proposition $P(n)$ later on has no meaning. So, be sure to *define* a statement if you want to refer to it!

  To be concise, you might say something like "Let $P(n)$ be the claim defined

above." (However, make sure that n is, indeed, the variable letter used in the claim above, for consistency's sake!)

- **Explicitly state that you are using mathematical induction and state the variable to which you are applying induction.**

  In the future, you might have an induction proof where multiple variable letters are floating around. Also, just because your overall proof follows some kind of inductive structure, do **not** necessarily expect a reader to just understand you are using induction. Telling them this information up front saves them a lot of trouble.

- **Be as *explicit* and thorough as possible in the Base Case.**

  Do *not* just write out what $P(1)$ means and expect the reader to understand why it is True. This onus is on you, the proof-writer!

  Do *not* just write out what the statement $P(1)$ is, itself, and put a ✓next to it. This does not prove anything!

  If the proposition $P(1)$ is some kind of equation (which is common), demonstrate why both sides are actually equal, instead of just writing the equation and expecting a reader to see why it works out.

- **The IH and IS together apply the *direct proof strategy* for $\implies$ statements.**

  The **IH** introduces an arbitrary and fixed natural number and assumes the left-hand side of the implication $P(k) \implies P(k+1)$. That's why it is our *hypothesis*. We then use this assumption to deduce $P(k+1)$. This proves the conditional statement in condition (2) in the PMI.

  Be sure to *quantify* the variable k here! A statement like "Assume $P(k)$" has no meaning. What is k? Is it a natural number?

  "Let $k \in \mathbb{N}$ and assume $P(k)$" is an acceptable form here. "Let $k \in \mathbb{N}$", to a mathematical reader, implicitly means "Let $k \in \mathbb{N}$ (be arbitrary and fixed)".

- **It helps to explicitly write out what $P(k)$ means in the IH.**

  For one, this helps a reader understand your assumption and follow the rest of the proof better.

  But also, this helps *you* figure out how to prove $P(k+1)$, which is your goal in this step. If you're struggling to work through this step in your head (perhaps on an exam or a homework problem), simply write out the meaning of $P(k)$ at the top of your paper and the meaning of $P(k+1)$ at the bottom. Now do you see how they might be connected? Try to work downwards from $P(k)$ and upwards from $P(k+1)$ and connect them in the middle.

- **You *must* invoke the IH somewhere in the Induction Step!**

  If you didn't use the **IH** at all, why did you need to use induction?

  When you use the **IH**, *say that you are doing so.* Don't expect the reader to remember/recognize that this is what you're doing.

- **Make a conclusion.**

  Tell the reader what you have accomplished.

Okay, now that we've dicussed how to write a good proof by induction, let's actually do so!

5.2.3 Examples

Here are a couple of examples of good induction proofs. Use them as guides when writing your own. We have omitted the usual discussion of how we came up with the arguments here, partly because we want to just emphasize the *structure* of the proofs, but also because we worked on those problem-solving aspects extensively in Chapter 2.

Notice that we are using abbreviations for some components of these proofs, namely **BC** (for Base Case) and **IH** (for Induction Hypothesis) and **IS** (for Induction Step). Feel free to use this shorthand, as well!

Example 5.2.5. *Sum of the odds is a square:*

Claim: The sum of the first n odd natural numbers is n^2.

(Note: We already saw this claim as a puzzle in Section 1.4.3, and then asked you to work through the inductive details in Section 2.3.4. We will present a good proof of the claim here.)

Proof. Let $P(n)$ be the proposition

$$\text{“}\quad 1+3+5+\cdots+2n-1=\sum_{i=1}^{n}(2i-1)=n^2\quad\text{”}$$

We will prove that $\forall n \in \mathbb{N}.\ P(n)$ by induction on n.

BC: Consider $n = 1$. Notice that

$$\sum_{i=1}^{1}(2i-1)=1 \quad \text{and} \quad 1^2=1$$

and so

$$\sum_{i=1}^{1}(2i-1)=1^2$$

Thus, $P(1)$ is True, because $1 = 1$.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $P(k)$ holds. This means

$$\sum_{i=1}^{n}(2k-1) = n^2$$

IS: Consider $k+1$. We can write

$$\sum_{i=1}^{k+1}(2i-1) = 2(k+1) - 1 + \sum_{i=1}^{k}(2i-1) = 2k + 1 + \sum_{i=1}^{k}(2i-1)$$

by separating out the $(k+1)$-th term of the summation.

We now use the **IH** to replace the summation on the right-hand side and deduce that

$$\sum_{i=1}^{k+1}(2i-1) = 2k + 1 + k^2$$

Factoring then tells us

$$\sum_{i=1}^{k+1}(2i-1) = (k+1)^2$$

and therefore $P(k+1)$ holds.

By the PMI, we conclude that $\forall n \in \mathbb{N}.\ P(n)$. □

Here's another good induction proof of a useful fact about geometric series.

Example 5.2.6. *Geometric series formula:*

Claim: For every $q \in \mathbb{R} - \{0, 1\}$ and for every $n \in \mathbb{N}$, the following formula holds:

$$\sum_{i=0}^{n-1} q^i = 1 + q + q^2 + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1}$$

Proof. Let $q \in \mathbb{R} - \{0, 1\}$ be arbitrary and fixed. Define $P(n)$ to be the proposition

$$\text{“ } \sum_{i=0}^{n-1} q^i = \frac{q^n - 1}{q - 1} \text{ ”}$$

We will prove $\forall n \in \mathbb{N}.\ P(n)$ by induction on n.

BC: Consider $n = 1$. Observe that

$$\sum_{i=0}^{n-1} q^i = \sum_{i=0}^{0} q^i = q^0 = 1$$

since $q \neq 0$. Also, observe that

$$\frac{q^n - 1}{q - 1} = \frac{q - 1}{q - 1} = 1$$

since $q \neq 1$. Thus, $P(1)$ holds.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed and suppose $P(k)$ holds. This means

$$\sum_{i=0}^{k-1} q^i = \frac{q^k - 1}{q - 1}$$

IS: WWTS $P(k+1)$ holds. (Remember, WWTS means "We want to show".) That is, WWTS

$$\sum_{i=0}^{k} q^i = \frac{q^{k+1} - 1}{q - 1}$$

noticing that $(k+1) - 1 = k$.

Observe that we can algebraically simplify and use our assumptions to write

$$\begin{aligned}
\sum_{i=0}^{k} q^i &= \left(\sum_{i=0}^{k-1} q^i\right) + q^k && \text{summation notation} \\
&= \frac{q^k - 1}{q - 1} + q^k && \text{invoking } \mathbf{IH} \\
&= \frac{q^k - 1 + q^k(q-1)}{q - 1} && \text{common denominator} \\
&= \frac{q^k - 1 + q^{k+1} - q^k}{q - 1} = \frac{q^{k+1} - 1}{q - 1} && \text{algebra}
\end{aligned}$$

This shows that $P(k+1)$ holds, as well.

By the PMI, $\forall n \in \mathbb{N}.\ P(n)$ holds. □

Follow-up question: Why did we need $q \notin \{0, 1\}$ in the claim?

What happens when $q = 0$? Where does this proof break down? Does the formula still holds? If so, prove it. If not, can you fix it?

Try answering the same questions for $q = 1$, as well.

5.2.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What does the PMI (Principle of Mathematical Induction) state? How is it proven?

(2) What is the Base Case of an induction proof? How does it relate to the statement of the PMI?

(3) How are the Induction Hypothesis and Induction Step of a proof related? How do they relate to the statement of the PMI?

(4) Why is it important to invoke the Induction Hypothesis somewhere in the Induction Step?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Prove that

$$\sum_{i=1}^{n} i^3 = \left[\frac{n(n+1)}{2}\right]^2$$

holds for every $n \in \mathbb{N}$.

(2) Prove that the square of every odd natural number is one more than a multiple of 8. That is, prove that

$$(2n+1)^2 - 1 \text{ is a multiple of } 8$$

for every $n \in \mathbb{N}$.

(3) Consider this claim: $7^n - 4^n$ is a multiple of 3, for every $n \in \mathbb{N}$.

Rewrite this claim using logical symbolic notation. Then, prove it by induction.

(4) Recall that the Fibonacci Numbers are defined by

$$f_0 = 0 \text{ and } f_1 = 1 \text{ and } \forall n \in \mathbb{N} - \{1\}\,.\; f_n = f_{n-1} + f_{n-2}$$

Prove the following claims hold for every $n \in \mathbb{N}$, by induction on n:

(a) $\displaystyle\sum_{i=1}^{n} f_i = f_{n+2} - 1$

(b) $\displaystyle\sum_{i=1}^{n} f_{2i-1} = f_{2n}$

(c) f_{4n} is a multiple of 3

(d) **Challenge 1:** (n is a multiple of 3) $\implies$ (f_n is even)

(e) **Challenge 2:** (n is *not* a multiple of 3) $\implies$ (f_n is odd)

5.3 Other Variants of Induction

Now that we're really comfortable with how induction works and have seen many examples, we can show you two modifications of this method. The idea is that there is "nothing special" about using induction to prove a statement holds for every $n \in \mathbb{N}$. Don't get us wrong; there is a *lot* that's special about $\mathbb{N}$! What we mean is that it's possible to use induction to prove that a statement holds for every $n \in S$, where S might be some other kind of set. We will describe these sets for you in the following discussions and examples.

5.3.1 Starting with a Base Case other than $n = 1$

We need to have a base case in an induction proof, but there's nothing that says it always has to be $n = 1$. Perhaps we have a proposition $P(n)$ that is True for $n = 1$ and $n = 2$, but then somehow False for $n = 3$ and $n = 4$, but then True for every n that is at least 5. How could we prove these claims? Well, we could just show the individual cases for $n = 1, 2, 3, 4$ separately, and then use induction to prove all the others. This will work because the set $\mathbb{N} - \{1, 2, 3, 4\}$ is also an *inductive set*. In terms of the Domino Analogy, this is like saying, "Let's just skip a few dominoes and start the line falling at $n = 5$. The rest will all fall down in the exact same way as we'd expect."

In fact, we can even allow ourselves to talk about *negative* integers here! Let's slide to the left on the number line a bit and imagine that we actually have a line of dominoes numbered starting from, say, -3. That is, we'd have Domino $\# -3$ and Domino $\# -2$ and Domino $\# -1$ and Domino #0 and Domino #1 and all the rest. We can start the line falling at $n = -3$ and know that they will all fall into each other in much the same way as before.

The whole idea here is that we still have an infinite line of dominoes moving off to the right with no gaps between them. It doesn't matter what numerical label we assign to the *first* domino. A line of dominoes like this will topple into each other no matter how we number that first one. This idea is what the next theorem encapsulates.

Theorem 5.3.1 (Induction with any base case)**.** *Let $P(n)$ be a variable proposition. Let $M \in \mathbb{Z}$ be arbitrary and fixed.*

Let $S = \{z \in \mathbb{Z} \mid z \geq M\}$.

Suppose that

(1) $P(M)$ holds True*, and*

(2) $\forall k \in S \bullet P(k) \implies P(k+1)$ holds True

Then $\forall n \in S \bullet P(n)$ holds True*.*

This theorem says exactly what we were discussing: if we want to prove a proposition holds for every value greater than or equal to some specific value (M, in the theorem statement), then we can just start inducting from that value.

We make that our **BC** and apply the **IH** and **IS** to every value greater than or equal to it. Everything else is exactly the same.

Formal Proof

For the sake of illustration and completeness, we will formally *prove* this theorem. We hope that the discussion above—referencing the Domino Analogy—will help you intuitively understand how this works. Working through this proof will not directly and immediately affect your ability to apply induction as a technique. However, we do think that reading through it and trying to understand *how* it works will give you a better grasp of induction and proof techniques, and it will perhaps give you a deeper apprecation of the mathematics at work here. Specifically, we will use PMI to prove this modified version of itself!

Proof. Let $P(n)$ be a variable proposition. Let $M \in \mathbb{Z}$ be arbitrary and fixed.

Let $S = \{z \in \mathbb{Z} \mid z \geq M\}$.

Suppose that

(1) $P(M)$ holds True, and

(2) $\forall k \in S \mathbf{.}\ P(k) \implies P(k+1)$ holds True

Our goal is to prove that $\forall n \in S \mathbf{.}\ P(n)$ holds True.

Define the proposition $Q(n)$ by setting

$$Q(n) \iff P(n + M - 1)$$

Notice that, by algebraically manipulating the inequality, we have

$$n \geq 1 \iff n + M - 1 \geq M$$

This means that our goal now is to prove that $\forall n \in \mathbb{N} \mathbf{.}\ Q(n)$ holds True.

(Doing so will prove that $\forall n \in S \mathbf{.}\ P(n)$.)

We will prove this by induction on n.

BC: We know $P(M)$ holds, by assumption. Notice that $n + M - 1 = M \iff n = 1$. This means $Q(1)$ holds.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $Q(k)$ holds.

IS: Since $Q(k)$ holds, we know $P(k + M - 1)$ holds.

Also, since $k \in \mathbb{N}$, we know $k \geq 1$. Thus, $k + M - 1 \geq M$.

Thus, By condition (2) that we assumed, we can deduce that $P((k+M-1)+1)$ holds, i.e. that $P(k + M)$ holds.

This tells us that $Q(k + 1)$ holds.

By PMI, we deduce that $\forall n \in \mathbb{N} \mathbf{.}\ Q(n)$ holds.

Then, by the definition of $Q(n)$, this tells us that $\forall n \in S \mathbf{.}\ P(n)$ holds. □

As we said, try to work through the details of the proof but, in general, just keep in mind the intuitive idea that we're just "sliding over" and using a different base case. The mechanics of the inductive process are identical.

Example

Let's see this modified technique in action. In fact, the example we will show you is of the flavor that we hinted at when introducing this method, wherein some proposition holds for a few small values, doesn't hold for some other small values, but does hold for every value after a certain point.

Example 5.3.2. *How* 2^n *compares to* n^2*:*

Claim:

$$2^n > n^2 \iff n \in \{0, 1\} \cup \{z \in \mathbb{N} \mid z \geq 5\}$$

That is, the *only* integers z that satisfy $2^z > z^2$ are $z = 0, 1, 5, 6, 7 \ldots$.

(We will leave it to you to play around and figure out how we might have come up with such a claim. Typically, as you will see in this section's exercises, such an inequality might be presented along with the question, "For which values of n does this hold?" In that case, you would have to do some scratch work to identify your claim before starting an induction proof.)

Proof. Let $P(n)$ be the proposition "$2^n > n^2$".

First, observe the following cases:

$2^0 > 0^2 \iff 1 > 0$	so $P(0)$ is True
$2^1 > 1^2 \iff 2 > 1$	so $P(1)$ is True
$2^2 > 2^2 \iff 4 > 4$	so $P(2)$ is False
$2^3 > 3^2 \iff 8 > 9$	so $P(3)$ is False
$2^4 > 4^2 \iff 16 > 16$	so $P(4)$ is False

Notice that whenever $z \leq -1$, we have $2^z < 1$ and $z^2 \geq 1$, so $2^z \not> z^2$. Thus, $P(n)$ is False for every n that satisfies $n \leq -1$.

Next, define S to be the set $S = \{z \in \mathbb{N} \mid z \geq 5\}$.

We will prove $\forall n \in S.\ P(n)$ holds by induction on n.

BC: Observe that $P(5)$ holds because $2^5 = 32$ and $5^2 = 25$ and $32 > 25$.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose that $P(k)$ holds.

IS: Since $k \in S$, we know $k \geq 5$ and so $k > 4$.

Thus, $k - 1 > 3$ and so $(k-1)^2 > 9$; certainly, then, $(k-1)^2 > 2$.

Observe the following chain of manipulations of this inequality:

$$\begin{aligned}(k-1)^2 > 2 &\implies (k-1)^2 - 2 > 0 \\ &\implies k^2 - 2k - 1 > 0 \\ &\implies k^2 > 2k + 1 \\ &\implies 2k^2 > k^2 + 2k + 1 \\ &\implies 2k^2 > (k+1)^2\end{aligned}$$

Since we observed the first inequality holds, we can deduce that the final inequality above also holds.

(Note: In case you didn't realize, this chain of reasoning is a solution to practice exercise 2 from Section 4.9.9! To work this out, we did some scratch work, starting with the desired inequality at the bottom and "working backwards" until we found something obviously True. In writing it up here, we started with that obvious fact and worked down to the desired conclusion.)

By the **IH** $P(k)$, we know $k^2 < 2^k$. This tells us

$$2k^2 < 2 \cdot 2^k = 2^{k+1}$$

Applying the transitivity property of inequalities, we can deduce that

$$(k+1)^2 < 2k^2 < 2^{k+1}$$

and so $P(k+1)$ holds.

By PMI, $\forall n \in S.\ P(n)$ holds.

Overall, we have considered every $z \in \mathbb{Z}$. We observed that $P(z)$ fails for $z \leq -1$, that it holds for $z = 0, 1$, that it fails for $z = 2, 3, 4$, and that it holds for $z \geq 5$. Together, these observations prove the claim. □

Phew! There was actually a lot going on in that proof. Did you notice that the claim was phrased as an *if and only if*, so that we had to consider *all* the integers in our proof? That was tricky, but we did it!

5.3.2 Inducting Backwards

This variant of induction is useful when a proposition $P(n)$ happens to hold for all values of n *less* than some particular value. In terms of the Domino Analogy, this is like imagining our infinite line of dominoes running off to the left, instead of to the right. We already know that, for the reasons discussed in the previous section, it doesn't matter how we number them. Now, we can see that it also doesn't matter which *direction* they're going; they'll obey the same principles! The following theorem encapsualtes this observation.

Theorem 5.3.3 (Backwards induction)**.** *Let $P(n)$ be a variable proposition. Let $M \in \mathbb{Z}$ be arbitrary and fixed.*

Let $S = \{z \in \mathbb{Z} \mid z \leq M\}$.

Suppose that

(1) $P(M)$ *holds* True*, and*

(2) $\forall k \in S \bullet P(k) \implies P(k-1)$ *holds* True

Then $\forall n \in S \bullet P(n)$ *holds* True*.*

Notice the differences between this and Theorem 5.3.1

Formal Proof

At this point in our development, we feel comfortable giving *you* important theorems to prove. Specifically, we want you to prove this modified version of PMI you see above, Theorem 5.3.3! Letting you work through the details yourself, instead of just seeing us perform them for you, will be far more helpful in the long run. Furthermore, the details of this proof we have in mind are quite similar to those for the proof we gave you (in Section 5.3.1) of Theorem 5.3.1.

Leaving a proof as an "exercise for the reader" is actually quite common in mathematics, and in mathematics books, in particular. We're just doing our part to help you get used to this phenomenon! ☺

Proof. Left for the reader as Exercise 1 in Section 5.3.4. □

We will not show an example of this method in action because we believe it is exactly like the standard method of induction we have already seen. In fact, if you worked through the details of the proof above, you can probably even see how to "make up" an example for this section by just modifying some examples we've already seen! (What if we reverse an inequality . . .)

5.3.3 Inducting on the Evens/Odds

Let's motivate this section with an observation, which will lead us into the first example usage of this method. Consider the sequence of perfect square numbers:

$$1,\ 4,\ 9,\ 16,\ 25,\ 36,\ 49,\ 64,\ 81,\ 100,\ 121,\ 144,\ \ldots$$

Look at what happens when we divide them by 8; specifically look at the *remainders* (indicated by the numerators of the fractions in each case):

$$0+\frac{1}{8},\ 0+\frac{4}{8},\ 1+\frac{1}{8},\ 2+\frac{0}{8},\ 3+\frac{1}{8},\ 4+\frac{2}{8},\ 6+\frac{1}{8},\ \ldots$$

Notice that we left fractions like $\frac{4}{8}$ and $\frac{2}{8}$ unsimplified, keeping the denominator as 8, to indicate the remainders. Those remainders follow this pattern:

$$1,\ 4,\ 1,\ 0,\ 1,\ 2,\ 1,\ \ldots$$

It looks like every other remainder is 1. In fact, it looks like the remainder is 1 when we divide an *odd* number squared by 8. Interesting! You might wonder whether this pattern continues. A reasonable way to address this idea is to just jump right in and try to prove this claim by induction and see if it works. If it does succeed, then we have successfully discovered and proven a fact. If it doesn't succeed, then we might be able to figure out *where* it fails and *why*. This is a good, general recommendation for mathematical discovery: if you want to see if something is True, just try to prove it and see what happens!

Example

Try to work through the details of this one on your own before reading on. In doing so, you will have to figure out how to induct on the set of *odd* natural numbers, not all the natural numbers as we have done before. We will actually present the proof of this claim and afterwards discuss how the method works, but you should absolutely work on this on your own first! ...

Example 5.3.4. *Remainders of odd squares when divided by 8:*

Claim: Let O be the set of odd natural numbers; that is,

$$O = \{n \in \mathbb{N} \mid \exists m \in \mathbb{N} \cup \{0\}\,.\; n = 2m + 1\}$$

Let $P(n)$ be the proposition "n^2 is 1 more than a multiple of 8". Then

$$\forall n \in O\,.\; P(n)$$

Proof. Let $P(n)$ be defined as in the claim above. We will prove $\forall n \in O\,.\; P(n)$ by induction on n.

BC: Observe that $1^2 = 1$ and $1 = 0 \cdot 8 + 1$ (i.e. 1 is a multiple of 8 plus 1). Thus, $P(1)$ holds.

IH: Let $k \in O$ be arbitrary and fixed. Suppose $P(k)$ holds.

IS: Our goal now is to deduce $P(k+2)$ holds. (This is because $k+2$ is the next odd natural number, after k.)

Since $k + 2$ is odd, by assumption, we know $\exists m \in \mathbb{N} \cup \{0\}\,.\; k = 2m + 1$. Let such an m be given.

By the **IH**, we know $\exists \ell \in \mathbb{N}\,.\; k^2 = 8\ell + 1$. Let such an ℓ be given.

Now, we take these observations and use them to see that

$$\begin{aligned}(k+2)^2 &= k^2 + 4k + 4\\ &= (8\ell + 1) + 4(2m + 1) + 4\\ &= 8\ell + 8m + 8 + 1\\ &= 8(\ell + m) + 1\end{aligned}$$

Since $\ell, m \in \mathbb{Z}$, we know $\ell + m \in \mathbb{Z}$, as well. Thus, $(k + 2)^2$ is one more than a multiple of 8. Therefore, $P(k + 2)$ holds.

By induction, $P(n)$ holds for every $n \in O$. □

Follow-up questions: Can you also prove that the remainders of *even* squares when divided by 8 are *not* 1? (This would make the claim an *if and only if* statement.) Can you identify a pattern in the remainders of those even squares? Can you prove your claims?

(*Hint:* You probably won't need induction for these claims!)

Discussion of Method

Let's discuss why this works. The underlying principle is exactly the same as the other forms of induction we have seen. The only difference lies in the induction step. Since the odd natural numbers are "two steps apart", our goal is to prove

$$\forall k \in O \bullet P(k) \implies P(k+2)$$

This encapsulates the same idea as standard induction: take one instance of the proposition and use it to deduce the "next" instance holds. The only difference here lies in what we mean by "next". For completeness' sake, we will state a theorem that conveys this method. Again, we will leave it to you to fill in the details of the proof.

Theorem 5.3.5 (Induction on the odds)**.** *Let O be the set of odd natural numbers.*

Let $P(n)$ be a variable proposition. Suppose that

(1) $P(1)$ holds, and

(2) $\forall k \in O \bullet P(k) \implies P(k+2)$

Then $\forall n \in O \bullet P(n)$ holds.

Proof. Left for the reader as Exercise 2 in Section 5.3.4. □

Thinking in a very similar way, we can see that induction on the *even* natural numbers will also work. Here is a theorem that states this. Again, we will leave the proof to you.

Theorem 5.3.6 (Induction on the evens)**.** *Let E be the set of even natural numbers.*

Let $P(n)$ be a variable proposition. Suppose that

(1) $P(2)$ holds, and

(2) $\forall k \in E \bullet P(k) \implies P(k+2)$

Then $\forall n \in E \bullet P(n)$ holds.

Proof. Left for the reader as Exercise 2 in Section 5.3.4. □

Combining and Modifying These Methods

Let's say we have a proposition $P(n)$ and we want to prove $P(n)$ holds for every $n \in \mathbb{N}$. Perhaps the proposition, and the underlying ideas, are somehow tricky, and a regular old induction proof completely eludes us. Maybe it's because of some algebraic trick, maybe we just can't see how to do it in the most efficient way, or maybe there's actually something profound underlying the proposition that prevents us from doing so. Whatever the reason, we might be able to use a combination of these new induction methods and prove the proposition holds for all $n \in \mathbb{N}$ in a couple of pieces.

We can think of these new methods as "jumping" induction methods. Proving a proposition holds for every odd natural number amounts to the exact same inductive technique as before, but we just "skip over" the evens by adjusting what happens in the induction step. The same goes for inducting on the evens (although we also adjust the base case slightly, since 2 is the first even, not 1). If we perform the "odds" method first and *then* the "evens" method, overall we have proved that the proposition holds for *all* naturals.

The following example does something just like this, but you'll notice that it actually makes "jumps" of size 3 (instead of 2, like with the "odds" and "evens" methods). We won't state and prove (or even ask you to do so) theorems that convey these methods. At this point, we will rely on our collective intuition for how induction works and point out that these theorems/proofs will be very similar to the ones we have been seeing. If you feel like getting the practice, or want to have them for your notes and records, by all means go ahead and state and prove theorems about the method we are about to use!

Example 5.3.7. *Powers of* 2 *and multiples of* 7*:*

Claim: For every $n \in \mathbb{N}$, the number $2^n + 1$ is *not* a multiple of 7.

(At this point, we recommend doing some scratch work to identify a pattern in the remainders of the numbers $2^n + 1$ when divided by 7. You'll see that they follow a *cycle* of length 3. Neato! That's essentially what we will prove here; it's just that the claim wasn't presented in that way, so we had to do some work on the side to reformulate it and come up with a proof.)

Proof. Define the sets A_1, A_2, A_3 to be

$$A_1 = \{n \in \mathbb{N} \mid \exists m \in \mathbb{N} \cup \{0\}\text{.}\ n = 3m + 1\} = \{1, 4, 7, 10, \dots\}$$
$$A_2 = \{n \in \mathbb{N} \mid \exists m \in \mathbb{N} \cup \{0\}\text{.}\ n = 3m + 2\} = \{2, 5, 8, 11, \dots\}$$
$$A_3 = \{n \in \mathbb{N} \mid \exists m \in \mathbb{N} \cup \{0\}\text{.}\ n = 3m\} = \{3, 6, 9, 12, \dots\}$$

(That is, these three sets partition $\mathbb{N}$ based on remainders when divided by 3.)

Let $P(n)$ be the proposition "$2^n + 1$ is not divisible by 3". We will prove $\forall n \in \mathbb{N}\text{.}\ P(n)$ holds, by induction.

Define the propositions $Q(n)$ and $R(n)$ and $S(n)$ as follows:

$$Q(n) \text{ is } \text{“ } \exists \ell \in \mathbb{N} \cup \{0\} \text{ . } 2^n + 1 = 7\ell + 3 \text{ ”}$$
$$R(n) \text{ is } \text{“ } \exists \ell \in \mathbb{N} \cup \{0\} \text{ . } 2^n + 1 = 7\ell + 5 \text{ ”}$$
$$S(n) \text{ is } \text{“ } \exists \ell \in \mathbb{N} \cup \{0\} \text{ . } 2^n + 1 = 7\ell + 2 \text{ ”}$$

Observe that

$$\forall n \in \mathbb{N} \text{ . } \big(Q(n) \vee R(n) \vee S(n) \big) \implies P(n)$$

This is because a number that is a multiple of 7 plus 3 is *not* a multiple of 7, and neither is a multiple of 7 plus 5 or a multiple of 7 plus 2.

First, we will prove that $\forall n \in A_1 \text{ . } Q(n)$ holds, by induction on n.

BC: Observe that $2^1 + 1 = 3 = 0 \cdot 7 + 3$. Thus, $Q(1)$ holds.

IH: Let $k \in A_1$ be arbitrary and fixed. Suppose $Q(k)$ holds.

IS: Our goal now is to deduce that $Q(k + 3)$ holds.

Since $k \in A_1$, we know $\exists m \in \mathbb{N} \text{ . } k = 3m + 1$. Let such an m be given.

By the **IH**, we know $\exists \ell \in \mathbb{N} \text{ . } 2^k + 1 = 7\ell + 3$. Let such an ℓ be given. This means $2^k = 7\ell + 2$.

We can deduce that

$$2^{k+3} = 2^3 \cdot 2^k = 8 \cdot (7\ell + 2) = 56\ell + 16$$

Thus,

$$2^{k+3} + 1 = 56\ell + 17 = 7(8\ell) + 14 + 3 = 7(8\ell + 2) + 3$$

and so $Q(k + 3)$ holds, as well. Therefore, $\forall n \in A_1 \text{ . } Q(n)$.

Second, we will prove that $\forall n \in A_2 \text{ . } R(n)$ holds, by induction on n.

BC: Observe that $2^2 + 1 = 5 = 0 \cdot 7 + 5$. Thus, $R(2)$ holds.

IH: Let $k \in A_2$ be arbitrary and fixed. Suppose $R(k)$ holds.

IS: Our goal now is to deduce that $R(k + 3)$ holds.

By the **IH**, we know $\exists \ell \in \mathbb{N} \text{ . } 2^k + 1 = 7\ell + 5$. Let such an ℓ be given. This means $2^k = 7\ell + 4$.

We can deduce that

$$2^{k+3} = 2^3 \cdot 2^k = 8 \cdot (7\ell + 4) = 56\ell + 32$$

Thus,

$$2^{k+3} + 1 = 56\ell + 33 = 7(8\ell) + 28 + 5 = 7(8\ell + 4) + 5$$

and so $R(k+3)$ holds, as well. Therefore, $\forall n \in A_2 \text{.}\ R(n)$.

Third, we will prove that $\forall n \in A_3 \text{.}\ S(n)$ holds, by induction on n.

BC: Observe that $2^3 + 1 = 9 = 1 \cdot 7 + 2$. Thus, $S(3)$ holds.

IH: Let $k \in A_3$ be arbitrary and fixed. Suppose $S(k)$ holds.

IS: Our goal now is to deduce that $S(k+3)$ holds.

By the **IH**, we know $\exists \ell \in \mathbb{N} \text{.}\ 2^k + 1 = 7\ell + 2$. Let such an ℓ be given. This means $2^k = 7\ell + 1$.

We can deduce that

$$2^{k+3} = 2^3 \cdot 2^k = 8 \cdot (7\ell + 1) = 56\ell + 8$$

Thus,

$$2^{k+3} + 1 = 56\ell + 9 = 7(8\ell) + 7 + 2 = 7(8\ell + 1) + 2$$

and so $S(k+3)$ holds, as well. Therefore, $\forall n \in A_3 \text{.}\ S(n)$.

Overall, we have proven that either $Q(n)$ or $R(n)$ or $S(n)$ holds for *every* natural number (depending on a number's remainder when divided by 3). Accordingly, every natural number has the property that $2^n + 1$ is *not* a multiple of 7. □

In actuality, we proved a *stronger* result in our proof than what the claim presented. That is, not only did we show that no number of the form $2^n + 1$ is a multiple of 7, but we also showed exactly *how* those numbers fail to be multiples of 7.

In this section's exercises, we have included a couple of exercises that guide you through a proof like this by identifying the "jumps" and the claims. In this chapter's exercises, Section 5.7, we have included some problems that might require this kind of argument (but we won't necessarily tell you the overall structure of the argument, as we've done here).

It's worth mentioning, at this point, that you could quite easily adapt these methods to any situation you face, as long as the "jumps" you want to make follow some easily identifiable *pattern*. In the previous example, we made jumps of size 3, and so we split the set of all natural numbers into three sets and jumped along within those sets. In essence, this relies on the fact that we had a "formula" for how to get to the "next" instance of the proposition: we start with $P(k)$ and try to deduce $P(k+3)$. You could conceivably make jumps of size 4, or 10, or even make jumps that *double* your value; that is, you could prove that some proposition $P(n)$ holds for every n that is a power of 2, say, by proving

$$P(1) \text{ holds, and } \forall n \in \mathbb{N} \text{.}\ P(n) \implies P(2n)$$

Again, all of this relies on having some kind of "formula" or "rule" that tells us what the *next* instance under consideration is. For this reason, *we cannot induct on the set of all prime numbers*. If you're trying to prove some fact holds

for every prime number, don't even bother trying to use induction! You'd have to have some "rule" that says, "If k is a prime number, then the *next* prime number is ...". If you know of such a rule, the world of mathematics would *love* to hear from you! This would answer many outstanding, unresolved questions about the prime numbers and make you the most famous mathematician in all of history. Seriously!

5.3.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How does the Domino Analogy describe an induction proof with a base case that is not 1?

(2) Write out a proof template for proving a proposition $P(n)$ that holds for every odd natural number greater than or equal to 7.

(3) Why can't we "induct on the primes"?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Prove Theorem 5.3.3.

(2) Prove Theorems 5.3.5 and 5.3.6.

(3) State a theorem that represents a method for inducting on the set of all multiples of 5. Prove your theorem.

(4) Consider the inequality $n^3 < 3^{n-1}$.

   (a) Prove that the inequality holds for every n that satisfies $n \geq 6$.

   (b) Prove that the inequality fails for $n \in \{1, 2, 3, 4, 5\}$. (This is easy.)

   (c) Prove that the inequality holds for every n that satisfies $n \leq 0$.

(5) Define a sequence of numbers by

$$x_1 = 2 \text{ and } x_2 = 2 \text{ and } \forall n \in \mathbb{N} - \{1, 2\}\,.\; x_n = x_{n-2} + 1$$

Let $P(n)$ be the proposition

$$\text{“ } x_n = \frac{1}{2}(n+1) + \frac{1}{4}\big(1 + (-1)^n\big)$$

(a) Let O be the set of odd natural numbers. Prove that $\forall n \in O\text{.}\ P(n)$ by induction.

(b) Let E be the set of even natural numbers. Prove that $\forall n \in E\text{.}\ P(n)$ by induction.

(6) Consider the following claim:

$$\sum_{k=1}^{n} (-1)^{k-1} k^2 = (-1)^{k-1} \sum_{k=1}^{n} k$$

That is, we claim that

$$1^2 - 2^2 + 3^2 - 4^2 + \cdots + (-1)^{n-1} n^2 = (-1)^{n-1}(1 + 2 + 3 + \cdots + 9)$$

holds for every $n \in \mathbb{N}$.

(a) Prove that the above formula holds for $n = 1$ and $n = 2$.

(b) Prove that whenever the formula holds for k, it also holds for $k + 2$.

(c) Explain intuitively why (a) and (b) prove the claim.

5.4 Strong Induction

Now, we will see why our previous work with induction constitutes “Regular” Induction. What follows is a technique known as **Strong Induction**. You will see why the term applies. Specifically, it refers to the inductive hypothesis, wherein we make a *stronger* assumption; informally, we will assume “more stuff” in that part of our proof, which allows us to make a conclusion more easily (or, sometimes, at all). The important part of this section, in addition to seeing several examples to get a handle on this modified technique, will be to actually *prove* that this stronger technique is even valid. To do that, we will actually invoke the PMI, itself!

5.4.1 Motivation

Look back to the examples from Section 2.4. There, we made some observations about the number of ways to tile a $2 \times n$ rectangular board with dominoes, and we played the game of Takeaway. In working through the inductive arguments for each of those examples, we found the situation to be slightly *different* than previous inductive arguments. When we proved something like

$$\sum_{k=1}^{n} \frac{n(n+1)}{2}$$

holds for every $n \in \mathbb{N}$, we could, in the inductive step, appeal to the immediately preceding case and invoke the inductive hypothesis, like so:

$$\sum_{k=1}^{n+1} k = (n+1) + \sum_{k=1}^{n} k = n + 1 + \frac{n(n+1)}{2} = \frac{(n+1)(n+2)}{2}$$

Of course, we didn't refer to these parts of our argument as the "**IH**" or "**IS**" yet, but that is what we were doing.

When we considered the Domino Tilings example, though, we found that we needed to refer to *two* previous instances of the fact. Specifically, to find the number of tilings of a $2 \times n$ board, we needed to know not only how many tilings of a $2 \times (n-1)$ board there were, but also how many tilings of a $2 \times (n-2)$ board there were. This is inherently different! What is it about an inductive argument that lets us do this? How does this follow the "domino analogy" we described? Or the "Mojo the Monkey" analogy? Does it, at all?

When we considered the game of Takeaway, we had even "more" different situation, didn't we? In constructing Player 2's winning strategy, we noticed that Player 2 should just mimic whatever Player 1 does, but on the other pile. That is, if Player 1 removes, say, 3 stones from the left pile, then Player 2 should remove 3 stones from the right pile, to guarantee a win. This held true no matter how many stones Player 1 removed. In that sense, we required the fact that Player 2 had a winning strategy on *any size* of piles up to n (inclusive) to guarantee that Player 2 had a winning strategy on piles of size $n+1$. This required a lot of assumptions to go into our inductive hypothesis. How do we know that we can do that?

5.4.2 Theorem Statement and Proof

Our goal now is to state and prove a modified version of the PMI that reflects these kinds of examples, the Domino Tilings and the Takeaway game. They represent inductive arguments where we might have to (1) refer to *more than one* previous instance to prove the subsequent instance of the claim, or (2) refer to *some unknown* previous instance to prove the subsequent instance. Both of those argument styles will be covered by this theorem. Let's see that statement first and then discuss what it means.

Theorem 5.4.1 (Principle of Strong Mathematical Induction (Strong PMI))**.** *Let $P(n)$ be a variable proposition. Suppose that*

(1) $P(1)$ holds True*, and*

(2) $\forall k \in \mathbb{N} \bullet \big(\forall i \in [k] \bullet P(i)\big) \implies P(k+1)$ holds True

Then $\forall n \in \mathbb{N} \bullet P(n)$ holds True*.*

Whoa, what does this say? We're giving you some extra work by presenting it here in logical notation before discussing it in a more wordy way, but we think you can handle it. Try to parse these two conditions, although surely condition

(2) is the tough one. What does it say? Read it out loud, write it down in an English sentence, think about it. Compare it to the regular old PMI we stated and proved in the previous section. Why would we call this one "strong"? How are these theorems different? Are their hypotheses different? What about their conclusions? Take a few minutes' break from reading to ponder these questions. Then, read on ...

Okay, let's explain this theorem. Notice that the *only* difference between The **Strong PMI** (Theorem 5.4.1) and The **Regular PMI** (Theorem 5.2.2) lies in condition (2), which governs what we do in the induction hypothesis part of a proof. The setup (that we have a variable proposition) and condition (1) (the base case) and the conclusion (that $P(n)$ holds for every $n \in \mathbb{N}$) are identical. Let's compare condition (2), now.

The Regular PMI requires that $P(k)$ is sufficient to allow us to deduce $P(k+1)$, for every $k \in \mathbb{N}$. If we can achieve that (the domino toppling affect), and we have a Base Case, then $P(n)$ holds for every $n \in \mathbb{N}$. This is what we do in the **IH** and **IS** of an induction proof: suppose $P(k)$ holds and use it to deduce $P(k+1)$ necessarily holds, too.

Let's rewrite condition (2) of The Strong PMI to see what it says:

$$\forall k \in \mathbb{N}\text{.}\ \big(\, P(1) \wedge P(2) \wedge P(3) \wedge \cdots \wedge P(k) \,\big) \implies P(k+1)$$

That is, Strong PMI requires that *all* of the previous instances of the proposition ($P(1)$ and $P(2)$ and $P(3)$ and ... and all the way until $P(k)$) are *together* sufficient to allow us to deduce $P(k+1)$. This theorem seems to say, "Hey, don't worry about using *just* $P(k)$ to get to $P(k+1)$; you can actually use *all* of the statements $P(1)$ through $P(k)$ to get there! The desired conclusion—that $\forall n \in \mathbb{N}\text{.}\ P(n)$—will still follow!" Isn't that nice?

There are three aspects of this theorem to discuss now: (1) why this method is actually valid, (2) when we would need to use it, and (3) how to use it. We can address aspect (3) quickly right now, before showing you some examples later on. The only difference between a Strong Induction proof and a Regular Induction proof will be in the **IH** and the **IS**. When using Strong Induction, in the **IH** we will suppose $P(1)$ and $P(2)$ and ... and $P(k)$ all hold, then use them to deduce $P(k+1)$ necessarily holds, too. In the **IS**, we will just have to be careful about pointing out *which* of the assumptions of the **IH** we use.

To address aspect (2)—when to use Strong Induction—we will show you several examples. In working through these examples, we will point out precisely why a regular induction proof would *fail*. By seeing several instances of this, we hope to develop some intuition for when to recognize these situations in the future. That is, we will learn to realize what kinds of claims *require* a strong **IH** in their proof.

Let's address aspect (1) right now, because it is the most pressing. Before we race on and start using a proof technique, we want to make sure it's actually mathematically valid! If you're like us, you're wondering, "How is this theorem

even True? It says that we need to know a whole lot more about how the instances of $P(n)$ relate to each other. Why should we be allowed to make so many assumptions in the **IH** and be able to use them later?"

A Modified Domino Analogy and a Heuristic Diagram

We'll start with a modification of the Domino Analogy from Chapter 2, and then show you a **heuristic diagram** for how Strong Induction works, to satisfy our intuitions. After that, we'll formally prove the theorem above.

Think about how Regular Induction followed the Domino Analogy. We only needed to know that Domino n will fall into Domino $n+1$ to guarantee the whole line will fall. Here, with Strong Induction, we actually need to know that *all* of the dominoes up to (and including) Domino n have fallen and knocked into Domino $n+1$, toppling it, to guarantee the whole line will fall. It's as if the dominoes are getting "heavier and heavier" as the line goes on, so we need a whole bunch of them falling into each other to generate enough momentum to topple the next, much heavier one.

Let's put this another way. Think about the chain of implications connecting all of our propositions. Our **BC** will tell us $P(1)$ is True. Great. This will imply that $P(2)$ holds. (Use $n = 1$ in condition (2) of SPMI.) Knowing these two will *together* imply that $P(3)$ holds. (Use $n = 2$ in condition (2) of SPMI.) Knowing all three of those will collectively imply $P(4)$ holds. And so and so on:

$$\underbrace{\underbrace{\overbrace{\big(P(1)\big)}^{\mathsf{T}\text{ by BC}}\overbrace{\Longrightarrow}^{\text{Use IS}}\overbrace{\big(P(2)\big)}^{\text{Get }\mathsf{T}}}_{\text{Know } P(1)\wedge P(2)}\underbrace{\Longrightarrow}_{\text{Use IS}}\underbrace{\big(P(3)\big)}_{\text{Get }\mathsf{T}}}_{\text{Know } P(1)\wedge P(2)\wedge P(3)} \implies \big(P(4)\big) \implies \big(P(5)\big) \implies$$

In some sense, this points to *why* the method works, overall. We prove $P(1)$ holds, just like with Regular Induction. But then, to "get to" the truth of $P(2)$, that first step—$P(1) \implies P(2)$—is the **same** in Strong Induction as it is in Regular Induction. (Use $n = 1$ in condition (2) of SPMI and PMI. It's the **same** condition.) From there on out, when we use Strong Induction, we're just making use of the fact that *all* of the previous instances of the proposition have held True; we might as well use them to keep traveling through and deducing the truths of the next propositions! Regular Induction doesn't bother with this. It says, "Okay, great, all of the previous instances have held. We don't actually need them to prove the next instance; we only need the immediately preceding one."

Here's one more slightly different way of interpreting this "chain of implications". This will actually directly hint at the proof we will see very shortly, as well! Pretend we're moving along with a Strong Induction process, and we've proven everything up until $P(n)$; that is, $P(1)$ and $P(2)$ and ... and $P(n)$ are all

True. Let's just package those instances all together and label them as one big proposition, $Q(n)$. (Thinking of it another way, we'll take all of those dominoes and bind them together into one giant domino.) The next step is to use that one instance to prove the next one, which sounds a lot more like Regular Induction, which we're more comfortable with for now. This is essentially what we will do in the proof! We'll reformulate the whole process of Strong Induction to phrase it as a *Regular Induction* process.

Formal Proof

As the previous paragraph hinted, the proof below will make use of PMI. (In fact, we will even use the proof template for a proof by induction that we saw in the previous section!) In this sense, we are really proving this statement:

$$\textbf{PMI} \implies \textbf{SPMI}$$

Let's do it!

Proof. Let $P(n)$ be a variable proposition. Suppose that

(1) $P(1)$ holds True, and

(2) $\forall k \in \mathbb{N}.\ \big(\forall i \in [k].\ P(i)\big) \implies P(k+1)$ holds True

Our goal is to prove that $\forall n \in \mathbb{N}.\ P(n)$.

Define the proposition $Q(n)$ by setting

$$Q(n) \iff \forall i \in [n].\ P(i)$$

(That is, $Q(n)$ says "All of the propositions $P(1)$ and $P(2)$ and ... and $P(n)$ are True.)

We will now prove $\forall n \in \mathbb{N}.\ Q(n)$ by induction on n.

BC: By the definition of the proposition $Q(1)$, we have $Q(1) \iff P(1)$. Condition (1) tells us that $P(1)$ holds, and therefore $Q(1)$ holds, as well.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $Q(k)$ holds.

IS: By the definition of $Q(k)$, we have

$$Q(k) \iff \forall i \in [k].\ P(i)$$

(Again, that is, $P(1)$ and ... and $P(k)$ all hold.)

By condition (2), we can deduce that $P(k+1)$ holds.

This means that $\forall i \in [k+1].\ P(i)$. (That is, we already knew $P(1)$ and ... and $P(k)$ all hold, and we just found that $P(k+1)$ holds, too.)

By the definition of $Q(k+1)$, this means that $Q(k+1)$ holds. This was the goal of the **IS**.

Accordingly, by PMI, we deduce that $\forall n \in \mathbb{N}.\ Q(n)$ holds.

By the definition of $Q(n)$, we have

$$\forall n \in \mathbb{N}.\ Q(n) \implies P(n)$$

(That is, every instance $Q(n)$ says that *all* of the instances $P(1)$ and ... and $P(n)$ hold and so, at the very least, we know that $P(n)$ itself holds.)

Since we just proved that $Q(n)$ holds for every $n \in \mathbb{N}$, we may deduce that $P(n)$ therefore *also* holds for every $n \in \mathbb{N}$, i.e.

$$\forall n \in \mathbb{N}.\ P(n)$$

This was the goal, so our proof is complete. □

Proof Summary and a Striking Equivalence

Look at what we've accomplished: we used Regular Induction to prove that Strong Induction is a valid technique. This tells us the PMI Theorem *implies* the SPMI Theorem, as we mentioned above:

$$\textbf{PMI} \implies \textbf{SPMI}$$

But certainly, this works the other way around, too! If we had already somehow proven (by other means) that Strong Induction is valid, then Regular Induction would have to be valid, as well. That is, we also know that

$$\textbf{SPMI} \implies \textbf{PMI}$$

Said another way: if we already had Strong Induction in hand as a valid proof technique, then whenever we would want to use Regular Induction to prove something, we would just use Strong Induction to accomplish our goal. In that sense, Strong Induction "subsumes" Regular Induction as a technique.

Together, these two observations tell us something remarkable about the theorems PMI and SPMI as they are in the world of mathematical truths. We have now shown that they are equivalent:

$$\textbf{PMI} \iff \textbf{SPMI}$$

Each theorem implies the other one.

Now, for the practical purpose of *applying* these techniques to prove facts, this equivalence might not seem to matter too much, but it really does tell us something helpful. It says this:

> *Whenever we have to prove something by induction, we might as well always use Strong Induction.*

Think about this for a few minutes. Read over the theorem statements and their proofs and consider it. Have it in mind as we work through the coming examples.

Once you've read the proof template below, go back to the examples from the previous section on Regular Induction and apply *Strong* Induction to them. Does it work? Does it seem different? Try it! We'll discuss this Regular/Strong comparison again after working out the examples below, so let's move on and see how to use Strong Induction.

5.4.3 Using Strong Induction: Proof Template

This template is very similar to the one for Regular Induction, since the only difference between the two theorems (and, accordingly, their respective techniques when applied) occurs in the **IH**.

Template for a "Proof by Strong Induction"

Goal: Prove that $\forall n \in \mathbb{N}.\ P(n)$

Proof.

Let $P(n)$ be the proposition "_____".

We will prove $\forall n \in \mathbb{N}.\ P(n)$ by induction on n.

Base Case: Observe that $P(1)$ holds because _____.

Induction Hypothesis: Let $k \in \mathbb{N}$ be arbitrary and fixed.

Suppose $\forall i \in [k].\ P(k)$ holds.

Induction Step: Deduce that $P(k+1)$ also holds.

By PMI, it follows that $\forall n \in \mathbb{N}.\ P(n)$. □

All of the same important observations and recommendations that we made about Regular Induction apply here, as well. We have to be sure to *define* a proposition, point out that we're using (strong) induction on a specific variable, label our steps, and make a conclusion.

One new recommendation we want to make is a refinement of an old one. When using Regular Induction, we had to be sure to cite *the* **IH** whenever we used it. Here, we will have *many* instances of the proposition in our **IH**, so we will actually have to be careful and cite *which* instance(s) of the proposition we use! You'll see this come into play in the examples below.

5.4.4 Examples

We will see three different "kinds" of examples here. Even though they all use the same template for Strong Induction that we just introduced, they differ in how the refer to the hypotheses in the **IH**. This first one is a direct application of the method, so let's work through it first, and then discuss how the other examples might be different.

Example 5.4.2. *A formula for a recursively-defined sequence:*

Claim: Let the sequence s_n be defined by

$$s_0 = 1 \text{ and } \forall n \in \mathbb{N}\text{.}\ s_n = 1 + \sum_{i=0}^{n-1} s_i$$

Find and prove a closed formula for s_n for every $n \in \mathbb{N} \cup \{0\}$.

Proof. Let $P(n)$ be "$s_n = 2^n$". We prove $\forall n \in \mathbb{N} \cup \{0\}\text{.}\ P(n)$ by induction on n.

BC: When $n = 0$, observe that $s_0 = 1$ and $2^0 = 1$, so $s_0 = 2^0$. Thus, $P(0)$ holds.

IH: Let $k \in \mathbb{N} \cup \{0\}$ be arbitrary and fixed. Suppose $P(0) \wedge P(1) \wedge \cdots \wedge P(k)$ holds.

IS: Observe that

$$\begin{aligned}
s_{k+1} &= 1 + \sum_{i=0}^{k} s_i && \text{Definition of } s_{k+1} \\
&= 1 + \sum_{i=0}^{k} 2^i && \text{Using } \textbf{IH}\text{s: } P(0) \wedge \cdots \wedge P(k) \\
&= 1 + \left(2^{k+1} - 1\right) && \text{Standard result (see Exercise 2.7.1)} \\
&= 2^{k+1}
\end{aligned}$$

Thus, $P(k+1)$ holds. Therefore, $\forall n \in \mathbb{N} \cup \{0\}\text{.}\ P(n)$ holds, by induction. □

Notice that this example required us to use *all* of the instances in the **IH**. Isn't that striking? Certainly, we *needed* strong induction here. Without knowing all of the previous instances held, we wouldn't have any hope of deducing the next one!

What distinguishes this from the next example is that here we knew exactly *which* instance(s) of the **IH** we used (namely, all of them). In the next example, we will invoke the **IH**, but we won't be able to say exactly which instance we use. You'll see what we mean!

Example 5.4.3. To start we need to introduce you to (or perhaps remind you of) a couple of ideas about prime numbers and the natural numbers.

Primes: A **prime number** is an element of the set

$$P = \{n \in \mathbb{N} \mid n > 1 \wedge (n = ab) \implies (a = 1 \vee a = n)\}$$

That is, the only divisors of a prime number are 1 and itself.

Prime Factorization: Given $x \in \mathbb{N}$, a **prime factorization** of x is a product of primes that equals x, with repeats allowed.

For example, a prime factorization of 6 is $2 \cdot 3$, and a prime factorization of 252 is $2 \cdot 2 \cdot 3 \cdot 3 \cdot 7$.

We will now state and prove the fact that every natural number has a prime factorization.

Claim: Let $F(n)$ be the proposition "n has a prime factorization". Then we claim that $\forall n \in \mathbb{N} - \{1\}.\ F(n)$.

Proof. We will prove $\forall n \in \mathbb{N} - \{1\}.\ F(n)$ by induction on n.

BC: Notice that $F(2)$ holds because $2 = 2$ is a prime factorization of 2.

IH: Let $k \in \mathbb{N} - \{1\}$ be arbitrary and fixed.

Suppose $\forall i \in [k] - \{1\}.\ F(i)$ holds. (That is, suppose $F(2) \wedge F(3) \wedge \cdots \wedge F(k)$.)

IS: Consider $k + 1$. We want to find a prime factorization of $k + 1$. There are two cases, based on whether $k + 1$, itself, is prime:

Case 1: If $k + 1$ itself is prime, then $k + 1$ is a prime factorization of $k + 1$, thereby showing $F(k + 1)$ holds.

Case 2: If $k + 1$ is not prime, there exist $a, b \in \mathbb{N} - \{1\}$ such that $k + 1 = a \cdot b$. Since $a, b \neq 1$, it must be that $1 < a < k + 1$ and $1 < b < k + 1$. That is, $2 \leq a \leq k$ and $2 \leq b \leq k$.

Thus, $F(a)$ and $F(b)$ hold, by the **IH**. Accordingly, there is a prime factorization of a and a prime factorization of b. Multiplying these two factorizations together yields a prime factorization of $a \cdot b = k + 1$. This shows $F(k + 1)$ holds.

In either case, we deduce that $F(k + 1)$ holds.

By induction, we conclude that $\forall n \in \mathbb{N} - \{1\}.\ F(n)$. □

Notice that we invoked the **IH** in this proof but we didn't know which "previous instance" of the claim we invoked. We were only able to appeal to *some* a and b with a certain property. This is different than the previous example, but it also clearly indicates we *needed* strong induction here. Nothing about a prime factorization for k could possibly help us find one for $k + 1$. Think about it: does knowing $14 = 2 \cdot 7$ help us figure out that $15 = 3 \cdot 5$? Does knowing $16 = 2^4$ help us figure out that 17 is prime?

This result we just proved is an important one: it says that every natural number has a prime factorization. Now, it also happens to be true that these prime factorizations are **unique**, in the sense that every natural number has *exactly one* prime factorization. Of course, this only works "up to the ordering of the factors". By this, we mean that $6 = 2 \cdot 3$ and $6 = 3 \cdot 2$ are really the *same* factorization of 6. Likewise, $252 = 2 \cdot 2 \cdot 3 \cdot 3 \cdot 7$ is the only factorization of 252; it is no different than writing $252 = 7 \cdot 3^2 codt 2^2$.

Nothing about proof above addresses this fact, though! We only used the *existence* of some a and b to deduce something. Who's to say that there weren't some *other* c and d with the same properties that we could have used? Think

about this. Can you prove that prime factorizations are unique? What method will you use?

The next example will adress the **Fibonacci Sequence**, a sequence of numbers we have used before. Specifically, we will state and prove a **closed form** for the sequence, which is typically defined recursively. By a "closed form" we mean a straight-forward expression that can be evaluated by "plugging and chugging". To find, for example, f_{100}, with the recursive definition of the sequence, we would have to compute *all* of the numbers in the sequence up until that point: we need f_{99} and f_{98}, which means we need f_{97}, which means ... With the closed form, though, we will be able to just "plug in n" and evaluate directly to find f_{100}.

Example 5.4.4. *A closed form for the Fibonacci Sequence:*

Claim: Define the standard Fibonacci Sequence by

$$f_0 = 0 \text{ and } f_1 = 1 \text{ and } \forall n \in \mathbb{N} - \{1\}\text{.}\ f_n = f_{n-1} + f_{n-2}$$

Define $\varphi = \frac{1+\sqrt{5}}{2}$. Then the following equality holds for every $n \in \mathbb{N} \cup \{0\}$:

$$f_n = \frac{1}{\sqrt{5}}\big(\,\varphi^n - (1-\varphi)^n\,\big)$$

Proof. Let f_n and φ be defined as in the claim above.

We will first prove the following equation:

$$1 + \varphi = \varphi^2 \qquad (\star_1)$$

Observe that

$$\varphi^2 = \left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = \frac{6+2\sqrt{5}}{4}$$
$$= \frac{3+\sqrt{5}}{2} = 1 + \frac{1+\sqrt{5}}{2} = 1 + \varphi$$

Then, we can use this to prove the following equation:

$$2 - \varphi = (1-\varphi)^2 \qquad (\star_2)$$

Observe that

$$(1-\varphi)^2 = 1 - 2\varphi + \varphi^2 = 1 - 2\varphi + (\varphi + 1) = 2 - \varphi$$

where we have used fact $(\star_1)$.

We will make use of both of these facts below.

Let $P(n)$ be the proposition

$$\text{“ } f_n = \frac{1}{\sqrt{5}}\big(\,\varphi^n - (1-\varphi)^n\,\big) \text{ ”}$$

We will prove that $\forall n \in \mathbb{N} \cup \{0\}\,.\ P(n)$ by induction on n.

BC: Observe that $f_0 = 0$ and

$$\frac{1}{\sqrt{5}}\big(\,\varphi^0 - (1-\varphi)^0\,\big) = \frac{1}{\sqrt{5}}(1-1) = 0$$

Thus, $P(0)$ holds.

IH: Let $k \in \mathbb{N} \cup \{0\}$ be arbitrary and fixed. Suppose $\forall i \in [k] \cup \{0\}\,.\ P(i)$ holds.

IS: Our goal now is to deduce that $P(k+1)$ holds.

Case 1: Suppose $k = 0$. Then we can directly observe that $f_1 = 1$ and

$$\frac{1}{\sqrt{5}}\big(\,\varphi^1 - (1-\varphi)^1\,\big) = \frac{1}{\sqrt{5}}\big(\,2\varphi - 1\,\big) = \frac{1}{\sqrt{5}}\big(\,1 + \sqrt{5} - 1\,\big) = \frac{1}{\sqrt{5}}\big(\,\sqrt{5}\,\big) = 1$$

This shows that $P(1)$ holds.

Case 2: Suppose $k \geq 1$. Then, observe that

$$\begin{aligned}
f_{k+1} &= f_k + f_{k-1} && \text{Defn, since } k \geq 1 \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^k - (1-\varphi)^k\,\big) + \frac{1}{\sqrt{5}}\big(\,\varphi^{k-1} - (1-\varphi)^{k-1}\,\big) && \textbf{IH}\text{s } P(k), P(k-1) \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^k + \varphi^{k-1} - (1-\varphi)^k - (1-\varphi)^{k-1}\,\big) && \text{Simplify} \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^{k-1}(\varphi+1) - (1-\varphi)^{k-1}((1-\varphi)+1)\,\big) && \text{Factor} \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^{k-1}\cdot\varphi^2 - (1-\varphi)^{k-1}(2-\varphi)\,\big) && \text{By } (\star_1) \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^{k+1} - (1-\varphi)^{k-1}(1-\varphi)^2\,\big) && \text{By } (\star_2) \\
&= \frac{1}{\sqrt{5}}\big(\,\varphi^{k+1} - (1-\varphi)^{k+1}\,\big)
\end{aligned}$$

Thus, $P(k+1)$ holds.

By induction, we conclude that $\forall n \in \mathbb{N} \cup \{0\}\,.\ P(n)$. □

A Discussion about Multiple Base Cases

Notice, in the previous example, that we had to establish two cases in the **IS**. Because the Fibonacci Sequence is defined recursively so that each term depends on two previous terms, we could not use the truth of $P(0)$ alone to deduce $P(1)$. We had to show $P(1)$ held *separately*. (Go back and try it. You'll find yourself trying to refer to f_{-1}, an undefined term!) After that, we can use the truth of $P(0)$ and $P(1)$ to deduce $P(2)$, then we can use $P(1)$ and $P(2)$ to deduce $P(3)$ ... That is to say, we really needed to throw in one *extra base case* before the whole "and so on" of induction kicked in.

There are two legitimate ways to handle this, and we just showed you one. The alternative would be to recognize that this situation will happen ahead of time, and instead present two base cases in the **BC** step. For the sake of illustration, let's show you how the relevant parts of the proof would be different, had we done that instead:

Proof.

...

...

BC: Observe that $f_0 = 0$ and

$$\frac{1}{\sqrt{5}}\big(\,\varphi^0 - (1-\varphi)^0\,\big) = \frac{1}{\sqrt{5}}(1-1) = 0$$

Thus, $P(0)$ holds.

Also, observe that $f_1 = 1$ and

$$\frac{1}{\sqrt{5}}\big(\,\varphi^1 - (1-\varphi)^1\,\big) = \frac{1}{\sqrt{5}}\big(\,2\varphi - 1\,\big) = \frac{1}{\sqrt{5}}\big(\,1+\sqrt{5}-1\,\big) = \frac{1}{\sqrt{5}}\big(\,\sqrt{5}\,\big) = 1$$

Thus, $P(1)$ holds.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $\forall i \in [k] \cup \{0\}\,.\ P(i)$ holds.

IS: Our goal now is to deduce that $P(k+1)$ holds. Observe that

...

... □

We moved the special $P(1)$ case *into* the **BC** section. Because of this, we had to also modify the *quantification* that happens in the **IH** and **IS**. We no longer want to use $k = 0$ in the ensuing argument, so in the **IH** we just take an arbitrary k that satisfies $k \geq 1$. However, we have already seen $P(0)$ holds, so we can still include it in our **IH**.

That's it! The two proofs are fundamentally identical. The only differences lie in their presentation and, even then, the differences are small. We will leave it to you to decide which style you prefer (if either) to be used in your proofs. We want to remind you, though, that these differences are small but they're also *subtle* and sometimes easy to forget! If you find yourself including many base cases, be sure to start your **IS** by seeking to prove a value above those base case values! You don't want to be inadvertently asserting some logical implication that doesn't actually hold. (For example, look back at the second proof above. If we had allowed $k = 0$ as a case in the **IS**, we would have inadvertenly been referring to f_{-1}, which does not exist. Thus, we would have been saying something incorrect, and the proof would be flawed, albeit not totally doomed.)

This kind of distinction typically occurs when you are asked to prove some representation formula for a recursively-defined sequence, where each term in the sequence is defined by several previous terms. There are several examples of this phenomenon in the exercises, both for this section and at the end of the chapter. Keep this in mind as you work on them!

Needing $n = 2$

A fairly common phenomenon that occurs in strong induction proofs is the necessity of proving *both* the $n = 1$ and $n = 2$ cases before jumping into the **IS**. In particular, this might happen when you have to prove some inequality or equality holds for n variables, where the $n = 1$ case is trivial and the $n = 2$ case is the interesting one that requires more work, and the rest of the induction follows by invoking the $n = 2$ case. Note that this requires taking $k \geq 2$ in the **IS**, of course.

Let's see an example to show you what we mean. Lucky for us, we have already proven the $n = 2$ case for this claim; it's actually one of DeMorgan's Laws for Sets!

Example 5.4.5. *A Generalized DeMorgan's Law for Sets:*

Claim: Let U be a universal set. For every $i \in \mathbb{N}$, let $A_i \subseteq U$ be a set.

Then, the following equality holds for every $n \in \mathbb{N}$:

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$$

Written another way, this claim says that, for every $n \in \mathbb{N}$, we have

$$\overline{A_1 \cup A_2 \cup \cdots \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}$$

Proof. Let U and $A_1, A_2, \ldots$ be defined as in the claim.

Let $P(n)$ be the proposition

$$\text{“} \overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i} \text{”}$$

We will prove $\forall n \in \mathbb{N}.\ P(n)$ by induction on n.

BC: Certainly, $\overline{A_1} = \overline{A_1}$, so $P(1)$ holds.

Also, $\overline{A_1 \cup A_2} = \overline{A_1} \cap \overline{A_2}$ by DeMorgan's Law for Sets (Theorem 4.6.9), so $P(2)$ holds.

IH: Let $k \in \mathbb{N} - \{1\}$ be arbitrary and fixed. Suppose $\forall i \in [k].\ P(i)$ holds.

IS: Our goal now is to deduce that $P(k+1)$ holds. First, observe that

$$\bigcup_{i=1}^{k+1} A_i = A_{k+1} \cup \bigcup_{i=1}^{k} A_i$$

so let's define

$$B_k = \bigcup_{i=1}^{k} A_i$$

Then, observe that

$$\begin{aligned}
\overline{\bigcup_{i=1}^{k+1} A_i} &= \overline{A_{k+1} \cup B_k} && \text{Defn of } B_k \\
&= \overline{A_{k+1}} \cap \overline{B_k} && \text{By } \mathbf{BC}, P(2) \text{ (a.k.a. DeMorgan)} \\
&= \overline{A_{k+1}} \cap \overline{\bigcup_{i=1}^{k} A_i} && \text{Defn of } B_k \\
&= \overline{A_{k=1}} \cap \bigcap_{i=1}^{k} \overline{A_i} && \text{By } \mathbf{IH}, P(k) \\
&= \bigcap_{i=1}^{k+1} \overline{A_i} && \text{Simplify}
\end{aligned}$$

Thus, $P(k+1)$ also holds.

By induction, $\forall n \in \mathbb{N}.\ P(n)$. □

5.4.5 Comparing "Regular" and Strong Induction

We want to reiterate something we said earlier when we introduced strong induction as a technique. It bears repeating here, because it's an important lesson:

> *Whenever we have to prove something by induction, we might as well always use Strong Induction.*

The reason behind this is that regular induction and strong induction are biconditionally connected; each one implies the other. When working through an induction proof, it essentially "doesn't hurt" to make a strong induction hypothesis, because we know we can. When you're working through a proof, you might not anticipate *which* or *how many* of the hypotheses from the **IH** that you'll need to invoke in the **IS**. It would be a shame to make a weaker hypothesis and find yourself referencing "truths" that you never officially proved! Instead, you might as well make the strongest hypothesis you can, just in case you'll need it. It might end up being overkill (in the sense that you really only needed $P(k)$ to deduce $P(k+1)$), but who cares, right? The point is to *prove* the fact at hand, and as long as that is achieved, then you've been successful.

As you move on in your mathematical careers, you'll probably get better at identifying the distinctions between regular/strong induction arguments. In particular, you'll likely notice when strong induction is truly *required*. Typically, this happens when we have a recursively-defined sequence, but this occurs in many other places, too. As you play around with a problem, trying to come

up with an argument, look at what sorts of *dependencies* there are between instances of your proposition. If you notice that an instance depends on several previous ones, you will almost certainly need a strong induction argument.

5.4.6 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What are the differences between Strong and Regular Induction?

(2) How might you identify when a Strong Induction argument is *required*?

(3) Why is it that we might as well always use Strong Induction, instead of deciding whether to use Regular/Strong?

(4) What was interesting about the use of the **IH** in the Prime Factorization example that we saw? How does it compare to the other examples, where we proved a formula about a recursively-defined sequence of numbers?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Define a sequence of numbers by

$$x_1 = 2 \text{ and } x_2 = 3 \text{ and } \forall n \in \mathbb{N} - \{1, 2\}\text{.}\ x_n = 3x_{n-1} - 2x_{n-2}$$

Prove that

$$\forall n \in \mathbb{N}\text{.}\ x_n = 2^{n-1} + 1$$

(2) Let the sequence a_n be defined by $a_0 = 0$ and $a_1 = 1$ and

$$\forall n \in \mathbb{N} - \{1\}\text{.}\ a_n = 5a_{n-1} - 6a_{n-2}$$

That is, $\langle a_n \rangle = \langle 0, 1, 5, 19, 65, 211, \ldots \rangle$

Prove that $a_n = 3^n - 2^n$ for every $n \in \mathbb{N} \cup \{0\}$.

(3) Let $a_1 \in \mathbb{Z}$ be arbitrary and fixed. Define a sequence by setting

$$\forall n \in \mathbb{N} - \{1\}\text{.}\ a_n = \sum_{k=1}^{n-1} k^2 a_k$$

Prove that

$$a_1 \text{ is even} \implies \forall n \in \mathbb{N}\text{.}\ a_n \text{ is even}$$

(4) Define a sequence $\langle t_n \rangle$ by setting

$$t_1 = t_2 = 2 \text{ and } \forall n \in \mathbb{N} - \{1, 2\}\text{.}\ t_n = \frac{1}{2t_{n-2}}(t_{n-1} - 4)(t_{n-1} - 6)$$

Prove that $\forall n \in \mathbb{N}\text{.}\ t_n = 2$.

(5) You have likely seen the **Triangle Inequality** before; it states

$$\forall x, y \in \mathbb{R}\text{.}\ |x + y| \leq |x| + |y|$$

(where $|x|$ is the *absolute value* of x). Prove that this inequality holds with n variables, not just 2; that is, prove that if we have a sequence of real numbers x_i, where $\forall i \in \mathbb{N}\text{.}\ x_i \in \mathbb{R}$, then

$$\forall n \in \mathbb{N}\text{.}\ \left|\sum_{k=1}^{n} x_i\right| \leq \sum_{k=1}^{n} |x_i|$$

(*Note:* Prove the $n = 2$ case, as well. We do not want you to just assume it!)

(6) Look back to Section 2.4.1 where we considered tiling $2 \times n$ rectangular chessboards with dominoes. Here, investigate the similar problem of tiling $3 \times n$ rectangular chessboards with straight triominoes (i.e. 3×1 rectangular pieces). Identify an inductive relationship and define a sequence that identifies the number of ways to tile a $3 \times n$ board.

(*Note:* Don't attempt to find a closed form and/or prove it! The techniques required to do so lie outside the scope of our current discussion. If you are curious, search for **recurrence relations**. If you'd like, try to apply what you read about to come up with a closed form for this problem. Can you prove it by induction?)

5.5 Variants of Strong Induction

Much like we saw a few variations on regular induction—inducting with a different base case, on a different set, backwards, etc.—we have a variation on strong induction to discuss here. As you will see, "minimal criminal" arguments are

essentially strong induction proofs where the induction step is done using a *contrapositive* conditional statement. These types of arguments pop up in induction proofs now and again, and understanding how they work will be very helpful!

Furthermore, we will state and prove a property of the set of natural numbers. This is known as the **Well-Ordering Principle**. Why is it included in this section? Well, you'll see how this principle is closely related to both induction and strong induction!

5.5.1 "Minimal Criminal" Arguments

Using a Contrapositive

Remember that a conditional statement is logically equivalent to its contrapositive. Also, remember that the theorem statements about how induction works all have conditional statements within them. They are always in condition (2), and they represent the action of the **IH** and **IS**. What happens if we consider the contrapositive of such a conditional statement? This won't change the truth of the theorem at all, but it will certainly affect how we apply induction as a proof technique. What would happen? Let's find out!

Here is the conditional statement from the Strong PMI:

$$\forall k \in \mathbb{N}.\ \big(\forall i \in [k].\ P(i)\big) \implies P(k+1)$$

Negating both sides and switching the arrow, we find its contrapositive:

$$\forall k \in \mathbb{N}.\ \neg P(k+1) \implies \big(\exists i \in [k].\ \neg P(i)\big)$$

When applying strong induction, we seek to prove that $P(1)$ and $\ldots$ and $P(k)$ together allow us to deduce $P(k+1)$. This new version of the statement reflects a different approach: supposing that $P(k+1)$ actually *fails*, let's deduce that *there is* a previous instance that also fails.

How It Works

Technically speaking, there's nothing new to say here! This method works because a conditional statement and its contrapositive are logically equivalent. However, this is a little unsatisfying. It feels funny to argue "backwards" like this, pretending that our proposition fails somewhere and showing that it also fails somewhere *earlier*. Isn't that the opposite of what we were trying to do? The crux of this method is twofold: (1) we already established a base case, and (2) this "earlier failure" argument is made for an *arbitrary* k.

Here's how we think of it. Say we have a proposition $P(n)$ and we want to show $\forall n \in \mathbb{N}.\ P(n)$. First, we establish that $P(1)$ holds. Good. Next, we pretend that $P(k+1)$ fails, for some *arbitrary* $k \in \mathbb{N}$. (Notice that $k+1 \geq 2$, so we aren't pretending that $P(1)$ fails, since we already know it holds.) We work through some argument and deduce that an *earlier* instance fails. Let's say $P(\ell)$ fails, for some ℓ that satisfies $1 \leq \ell \leq k$.

Now, this argument we just made was for an *arbitrary* k, so the same argument applies to this new value ℓ that we've produced. This guarantees that $P(m)$ fails, for some m that satisfies $1 \leq m \leq \ell - 1$. Then, the same argument can be re-packaged to apply to the value of m, and then ... You might see where this is going. Eventually, we "run out" of previous instances at which the proposition could fail; we *have* to eventually get back to $P(1)$. But we already know $P(1)$ holds!

The main idea can be sumamrized thusly: if we have a valid base case, and there is **no smallest instance that fails**, then the proposition holds everywhere. This is where the phrase "Minimal Criminal" comes from. (It is chosen for both its descriptiveness and its playful slant rhyme, of course.) "Criminal" refers to an instance where the proposition *fails*, and proving the implication

$$\forall k \in \mathbb{N}\,.\, \neg P(k+1) \implies \big(\exists i \in [k]\,.\, \neg P(i)\big)$$

amounts to showing that there can be no "minimal" such instance.

Another phrase that encapsulates this same idea is "No Least Counterexample". You might find this phrase in other books, so be aware that it refers to the same idea. It conveys the idea that there is no counterexample to the claim such that all the previous instances are true. Also, another term for this method is "Infinite Descent". It's less immediately clear that this refers to the same concept, because it is hinting at the actual *description* we gave of how this method works. By proving that we can always find a smaller counterexample, we are showing that there exists a "backwards" sequence of instances where our proposition fails. However, this sequence cannot be an "infinite descent" because we'll eventually run into $P(1)$, which we proved to be valid. Be aware that both of these terms are also used. We chose "Minimal Criminal" because it's more fun to say.

Proof Template

Let's briefly show you a template for how to write a proof like this, and then we'll move right into working through an example proof of an interesting fact. There isn't anything particularly new right here. We're applying the direct proof strategy to a $\implies$ statement; it's just that this statement is the contrapositive of a statement we've already seen before.

Template for a "Proof by a Minimal Criminal Argument"

Goal: Prove that $\forall n \in \mathbb{N}\,.\, P(n)$

Proof.

Let $P(n)$ be the proposition "_____".

We will prove $\forall n \in \mathbb{N}\,.\, P(n)$ by induction on n (a "minimal criminal"

argument, in fact).

Base Case: Observe that $P(1)$ holds because _____ .

Induction Hypothesis: Let $k \in \mathbb{N}$ be arbitrary and fixed.

Suppose $P(k+1)$ is False.

Induction Step: Deduce that $\exists \ell \in \mathbb{N}$ that satisfies $1 \leq \ell \leq k$ and such that $P(\ell)$ is False.

It follows that $\forall n \in \mathbb{N}.\ P(n)$. □

If you're worried about forgetting the technical details of this template, just keep the main idea in your mind:

> A "Minimal Criminal" argument works by applying the **contrapositive** of the usual **IH** and **IS** steps of an induction proof.

Example

The following result is interesting in its own right. (In fact, we will use it later on in Section 7.6.3 when we talk about how "big" infinite sets are. Neat, right?) We encourage you to play around with the claim first before jumping into the proof. Try to see why it's true and how it works. Check it for small values of n. Then, as you read the proof, look at your scratch work and see how it mimics the kinds of patterns you might have observed.

Example 5.5.1. *Expressing naturals uniquely as a product:*

Claim: Every $n \in \mathbb{N}$ can be expressed *uniquely* as a power of 2 times an odd number. That is,

$$\forall n \in \mathbb{N}.\ \exists m, \ell \in \mathbb{N} \cup \{0\}.\ n = 2^m \cdot (2\ell + 1)$$

and the ℓ, m that exist are the *only* values that satisfy this equality.

Proof. We prove this claim by induction on n; specifically, we use a "minimal criminal" argument.

BC: Observe that $n = 1$ has such a representation as $1 = 2^0 \cdot (2 \cdot 0 + 1)$. Furthermore, this is the *only* such representation because any other power of 2 will make the product at least 2, and any other odd will make the product at least 3.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose that $P(k+1)$ fails, i.e. that $k+1$ has no such representation, or it has more than one such representation. We will have two cases based on the parity of $k+1$.

Case 1: Suppose $k+1$ is even. This means $\frac{k+1}{2} \in \mathbb{N}$.

First, suppose $k+1$ has *no* such representation. Then, neither does $\frac{k+1}{2}$; for if

it actually *did*, then we could simply double it (i.e. increasing the power of 2 by 1) to find a representation of $k+1$.

Thus, $P\left(\frac{k+1}{2}\right)$ fails in this case (for non-existence).

Next, suppose $k+1$ has at least *two* such representations:

$$k+1 = 2^{m_1}(2\ell_1+1) \text{ and } k+1 = 2^{m_2}(2\ell_2+1)$$

We are assuming they are different, i.e. $(m_1, \ell_1) \neq (m_2, \ell_2)$. Since $k+1$ is even, we know $m_1, m_2 \geq 1$. By descreasing the powers of 2 by 1 each, we see that

$$\frac{k+1}{2} = 2^{m_1-1}(2\ell_1+1) \text{ and } \frac{k+1}{2} = 2^{m_2-1}(2\ell_2+1)$$

are two representations of $\frac{k+1}{2}$. (Also note that $m_1 - 1, m_2 - 1 \geq 0$.) These are *different* because $(m_1 - 1, \ell_1) \neq (m_2 - 1, \ell_2)$, based on our assumption above.

Thus, $P\left(\frac{k+1}{2}\right)$ fails in this case (for non-uniqueness).

In either situation, we find that $P\left(\frac{k+1}{2}\right)$ fails.

Case 2: Suppose $k+1$ is odd. This means $\exists \ell \in \mathbb{N} \cup \{0\}\,.\; k+1 = 2\ell+1$. Then certainly we can represent $k+1$ as

$$k+1 = 2^0 \cdot (2\ell+1)$$

Also, there is certainly no *other* way to do this. Using a different power of 2 will make the product even (but $k+1$ is odd), and using a different odd factor will make the product different. Thus, this case is a contradiction. ※

By indcution, then, $\forall n \in \mathbb{N}\,.\; P(n)$ holds. □

Interesting, isn't it? There was actually a little bit more going on, logically speaking, in this proof than we led on at the beginning. Specifically, the cases based on parity make this a little trickier. One of the cases (the even one) follows the "minimal criminal" argument. The other case (the odd one) can actually truly be proven. In this proof, we assumed $P(k+1)$ *failed*, but then realized it actually couldn't when $k+1$ is odd. That was the contradiction. It seems a little roundabout in retrospect, but it allows us to present the entire proof as a "minimal criminal" argument, rather than just doing two separate proofs, one for odds and one for evens.

Furthermore, we had to address not only the existence but the *uniqueness* of these representations. This is why there were two considerations to make in the case for $k+1$ being even. To show existence of these representations, we had to show it is not possible for $k+1$ to have *zero* representations; to show uniqueness, we had to show it is not possible for $k+1$ to have *two* representations.

5.5.2 The Well-Ordering Principle of $\mathbb{N}$

Motivation

We are all familiar with the relationship "$\leq$" on the natural numbers $\mathbb{N}$. Given any two elements $x, y \in \mathbb{N}$, it must be that one of the two relationships holds: either $x \leq y$ or $y \leq x$ (or possibly both, but only when $x = y$). We also know that

$$\forall x, y, z \in \mathbb{N}\,\boldsymbol{.}\ \big(x \leq y \wedge y \leq z\big) \implies x \leq z$$

and that $\forall x \in \mathbb{N}\,\boldsymbol{.}\ x \leq x$. This makes $\mathbb{N}$ an **ordered** set; we call "$\leq$" an *order relation* on $\mathbb{N}$. (See Section 6.3 for more information.)

Furthermore, it turns out that this relationship is a **well-ordering**. We won't define this term formally, but one of the key aspects of being a well-ordering is not having any *infinitely-descending chains*. Just think about how this works in $\mathbb{N}$: Does there exist an *infinite* sequence of elements $a_1, a_2, a_3, \ldots$ such that $a_1 > a_2 > a_3 > \cdots$? Is that possible? (Notice these inequalities are *strict*). No, it's not! The idea is that, starting from some number $a_1 \in \mathbb{N}$, if we "descend" we have to eventually reach 1. We can't "go lower" than that.

Rather than discuss well-orderings in generality—which you can do in a class on set theory or formal logic—we will just discuss how this concept works in the context of $\mathbb{N}$. It's a useful property, and we will have occasion to cite it in the future, too! In this section, we will state the principle, get your help in proving it, and then demonstrate its relationship with induction.

Statement and Proof

Theorem 5.5.2. *Every non-empty subset of* $\mathbb{N}$ *has a* least element*. Stated in logical form,*

$$\forall S \in \mathcal{P}(\mathbb{N})\,\boldsymbol{.}\ [S \neq \varnothing \implies (\exists \ell \in S\,\boldsymbol{.}\ (\forall x \in S\,\boldsymbol{.}\ \ell \leq x))]$$

Think about how this relates to our statement before that we cannot have an infinitely-descending chain of natural numbers. *If* we did have such a chain, we could define S to be the set of all of those elements in the chain. This set would **not** have a least element. Given an element of that set, we know it is one of the elements in the chain; let's say it's a_n. Then, a_{n+1} is also in the set and $a_{n+1} < a_n$. Thus, there would be no least element.

We will have you prove this theorem, because we think it will be instructive to work through the details. It is outlined for you in several steps. One key observation is that the proof is **by induction**! That is, by proving the Well-Ordering Principle in this way, we will have shown that the Principle of Mathematical Induction *implies* the Well-Ordering Principle.

Proof. By induction. Left for the reader as Exercise 5.7.21 □

An easy extra observation to make is that the least element of any subset $S \subseteq \mathbb{N}$ must also be *unique*. That is, we can't have two (or more) least elements. Let's say we actually do have two least elements of a set S; call them ℓ and m.

By the definition of what it means to be the least element, we would know $\ell \leq m$ and $m \leq \ell$. Of course, this tells us $\ell = m$, so they're the same!

Induction, Strong Induction, and The WOP

As we mentioned before, because we used induction to prove the Well-Ordering Principle, this shows the Principle of Mathematical Induction *implies* the Well-Ordering Principle. The next theorem shows that, in fact, those two theorems are **logically equivalent**: they imply *each other*. Furthermore, it says the the Principle of *Strong* Induction also implies the Well-Ordering Principle, and vice-versa. In fact, it says that all three of these theorems are logically equivalent!

Essentially, this is saying that these three theorems

Theorem 5.5.3. *The following are all logically equivalent:*

- *The Principle of Mathematical Induction*
- *The Principle of Strong Induction*
- *The Well-Ordering Principle*

Proof. Let's use the following shorthand for each theorem:

- **PMI**: The Principle of Mathematical Induction
- **PSI**: The Principle of Strong Induction
- **WOP**: The Well-Ordering Principle

By the way we proved PSI and WOP, we can already deduce that

$$\text{PMI} \implies \text{PSI} \quad \text{and} \quad \text{PMI} \implies \text{WOP}$$

We also described in Section 5.4.2 how

$$\text{PSI} \implies \text{PMI}$$

so now we know that

$$\text{PMI} \iff \text{PSI} \quad \text{and} \quad \text{PMI} \implies \text{WOP}$$

To complete the proof, we will show that WOP $\implies$ PMI. This will show that WOP$\iff$PMI, and the equivalence above will allow us to deduce that all three are logically equivalent.

To prove this, let's assume that the WOP is valid. We will use it to prove PMI. (Look back at the statement of PMI, in Theorem 5.2.2, to remind yourself why what we're about to do accomplishes our goal.)

Suppose we have a proposition $P(n)$, defined for every $n \in \mathbb{N}$. Let's suppose that $P(1)$ is True, and that $\forall k \in \mathbb{N}\text{.}\ P(k) \implies P(k+1)$. We need to show that $\forall n \in \mathbb{N}\text{.}\ P(n)$ holds.

Define the set F to be the set of "False instances" of $P(n)$. That is, define

$$F = \{n \in \mathbb{N} \mid P(n) \text{is False}\}$$

To prove that $\forall n \in \mathbb{N}\,.\ P(n)$, we will AFSOC that $F \neq \varnothing$.

Notice that $F \subseteq \mathbb{N}$, because we used set-builder notation. By our assumption in the line above, $\exists f \in F$. Let such an f be given.

Because of these two conditions, the WOP applies to the set F, telling us that F has a least element. Let ℓ be that least element. We know that $\ell \in F$ and,

$$\forall x \in F\,.\ \ell \leq x$$

Consider the case that $\ell = 1$. This is not possible, because our assumption above says that $P(1)$ holds, so $1 \notin F$.

Now, consider the case that $\ell \geq 2$. Our assumption above said that

$$\forall k \in \mathbb{N}\,.\ P(k) \implies P(k+1)$$

which is logically equivalent to

$$\forall k \in \mathbb{N} - \{1\}\,.\ \neg P(k) \implies \neg P(k-1)$$

by taking the contrapositive.

Applying this to the element $\ell \in \mathbb{N} - \{1\}$, we deduce that $\neg P(\ell - 1)$ also holds. That is, $P(\ell - 1)$ is False.

This means that $\ell - 1 \in F$. However, this contradicts the fact that ℓ is the **least** element of F, since $\ell - 1 < \ell$. ⨳

Therefore, it must be that $F = \varnothing$, which means that $\forall n \in \mathbb{N}\,.\ P(n)$.

This shows that the theorem PMI is valid. □

Look at the main part of this proof. To prove that $P(n)$ holds for *all* n, we supposed it failed for a *particular* n, the element $f \in F$. From there, you might be tempted to say, "Well, $P(f)$ failing means $P(f-1)$ fails, which then means $P(f-2)$ also fails, ... **and so on**, all the way down to $P(1)$, but we know that $P(1)$ is True." But that argument about "and so on" is *precisely* what PMI and WOP are all about! You can't use a hand-wavey "just keep going" argument to prove the very idea that you're allowed to make such arguments! This is why we invoked the WOP to produce the *least* element of F. It might have seemed odd to you that we would introduce $f \in F$ and the never use it again. We needed it to exist to guarantee $F \neq \varnothing$, which allows us to *apply* WOP.

5.5.3 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you

can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between a "Minimal Criminal" argument and a Strong Induction proof?

(2) We proved that every $n \in \mathbb{N}$ can be written as a product of a power of 2 and an odd number. What *else* is true about this representation?

(3) We proved that $\mathbb{N}$ is well-ordered. Do you think $\mathbb{Z}$ also has the property? What about $\mathbb{Q}$? What about $\mathbb{R}$?

(4) What does it mean that PMI and PSI and WOP are all *logically equivalent*?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Prove the Well-Ordering Principle. It's Exercise 5.7.21. Seriously, do it!

(2) Prove that $\sqrt{3}$ is irrational.

(**Hint:** AFSOC that $\sqrt{3} = \frac{a}{b}$, where $a, b \in \mathbb{N}$ and the fraction is in *reduced form*. Use a descent argument to contradict this *reduced form* assumption.)

(3) Use the Well-Ordering Principle to prove that every natural number, except 1, can be written as a sum of non-negative multiples of 2 and 3.

For example, $2 = 2$ and $8 = 6 + 2$ and $101 = 3 \cdot 33 + 2$.

(4) Consider the following equation: $4x^4 + 2y^4 = z^4$. In this problem, you will prove that this equation has **no** solution, $(x, y, z) \in \mathbb{N}^3$, by an argument that appeals to the Well-Ordering Principle.

(a) AFSOC $(x, y, z) \in \mathbb{N}^3$ *is* a solution, and suppose further that this solution has the *smallest* value of x amongst all solutions.

That is, we are defining

$$T = \{x \in \mathbb{N} \mid \exists y, z \in \mathbb{N}\text{.}\ 4x^4 + 2y^4 = z^4\}$$

and *pre-supposing* that this set is non-empty (i.e. the equation has solutions) so that T has a least element.

(b) Deduce that z is even.

Hint: In this and the next two parts, you may use the fact that a sum/difference of multiples of some natural number m is *also* a multiple of m.)

(c) Deduce that y is even.

(d) Deduce that x is even.

(e) Use this to deduce that there is another solution (a, b, c) with a *smaller* value of the first variable, i.e. $a < x$.

(f) Explain why this has proven that there are no solutions.

5.6 Summary

Now, we have finally placed **induction** on solid, mathematical ground! We have been building towards this for a while, so we wanted to present this fully when we finally got here. We took care to formally state *and* prove the Principle of Mathematical Induction, and see several examples of it in action. Then, we used PMI to prove the more general Principle of **Strong** Induction. In so doing, we pointed out that any induction proof *might as well* be a strong induction one, because one technique "subsumes" the other. Furthermore, we later proved—in the section about the Well-Ordering Principle of $\mathbb{N}$—that the two principles of induction we introduced are *logically equivalent* to each other (and to the Well-Ordering Principle, as well).

We saw a few variants of induction and an example or two for each one. One of the more helpful techniques we will use later on is the "minimal criminal" argument. This amounts to an induction proof where the induction step proves the *contrapositive* of the conditional statement required.

For all of these variants of induction, we provided you with some proof templates. Consult them in the future, and use them to make your proofs well-organized, clear, and easy to read. Not only will this make it easier for a reader to understand your written work, it will also reiterate the important concepts behind these proof techniques. These are not created by us out of pedantry, mind you: they are based firmly on the underlying principles!

The exercises below will give you lots of practice in working with all kinds of inductive arguments. We have posed some problems that are significantly more challenging than the ones we saw in Chapter 2. This is because we have now thoroughly studied the principle of induction and feel confident applying it to solve problems. Furthermore, some of the results you prove in these problems are interesting and helpful facts to have at our disposal. We will have occasion to refer to some of them in our later work in this book, even!

5.7 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them.

Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ⋆.

Problem 5.7.1. Prove that

$$\forall n \in \mathbb{N}. \sum_{k=1}^{n} k^3 = \left(\sum_{k=1}^{n} k \right)^2$$

Problem 5.7.2. For each of the following inequalities, determine the set of *natural numbers* for which it holds. Make a **claim** and then **prove** it (by induction, if necessary).

(a) $3^n > n^4$

(b) $(n-3)^2 > (n-2)^3$

(c) $3^n < n!$

(d) $4^n > n^4$

Problem 5.7.3. What is wrong with the "proof" of the following claim?

Claim: Every even natural number is a power of 2.

> We prove this by induction on n.
>
> Notice that $2 = 2^1$ is a power of 2.
>
> Next, suppose $k \in \mathbb{N}$ and $k \geq 4$ and k is even. Suppose that every even natural number up to (but not including) k is a power of 2.
>
> Since k is even, we can consider $\frac{k}{2}$. By assumption, $\frac{k}{2}$ is a power of 2, so $\frac{k}{2} = 2^j$ for some j.
>
> This shows that $k = 2^{j+1}$, so k is a power of 2.

Problem 5.7.4. Let's say that a number $n \in \mathbb{N}$ is "Special in the Land of (x, y)" if n can be written as a sum of non-negative multiples of x and y.

For example, 11 is Special in the Land of $(3, 5)$ because $11 = 5 + 2 \cdot 3$. Also, 15 is Special in that Land because $15 = 3 \cdot 5 + 0 \cdot 3$. However, 7 is not Special there.

For each of the following pairs (x, y), state and prove a claim that identifies the set $S_{x,y}$ of all numbers that are Special in the respective Land.

1. $(2, 3)$
2. $(3, 5)$
3. $(4, 9)$
4. $(7, 6)$

Problem 5.7.5. Prove that for any $n \in \mathbb{N}$, for any real numbers $x_1, x_2, \ldots, x_n$ that satisfy $\forall i \in [n]\,.\ 0 \leq x_i \leq 1$, the following inequality holds:

$$\prod_{i=1}^{n}(1 - x_i) \geq 1 - \sum_{i=1}^{n} x_i$$

This is known as **Bernoulli's Inequality**.

Problem 5.7.6. Let $P(n)$ be a proposition that depends on the variable n, which is allowed to assume any **integer** value.

For each of the following situations, you are given some kind of "Base Cases" and some kind of "Inductive Implication". Identify and explain which instances of the proposition you could **necessarily** deduce, from those assumptions.

For instance, if you were given $P(3)$ as a Base Case and $\forall n \in \mathbb{N}\,.\ P(n) \implies P(n+1)$ as an Inductive Implication, a correct answer would be "We would know $P(n)$ holds for every $n \in \mathbb{N}$ with $n \geq 3$."

1. Base Cases: $P(-3)$. Implication: $\forall n \in \mathbb{Z}\,.\ P(n) \implies P(n+1)$
2. Base Cases: $P(1) \wedge P(2)$. Implication: $\forall n \in \mathbb{N}\,.\ P(n) \implies P(2n)$
3. Base Cases: $P(0)$. Implication: $\forall n \in \mathbb{Z}\,.\ P(n) \implies \big(P(n-1) \wedge P(n+1)\big)$
4. Base Cases: $P(-1) \wedge P(0)$. Implication: $\forall n \in \mathbb{N}\,.\ P(n) \implies P(n+2)$.

Problem 5.7.7. Prove that for any integers $x, y \in \mathbb{Z}$ (with $x \neq y$), the number $x^n - y^n$ is a multiple of $x - y$, for every $n \in \mathbb{N} \cup \{0\}$.

Problem 5.7.8. (a) Identify the set of natural numbers n for which the inequality $n! > 2^n$ holds.

(Recall: $n! = n \cdot (n-1) \cdots 1$.)

(b) Identify the set of natural numbers n for which the inequality $n! > 3^n$ holds.

(c) Identify the set of natural numbers n for which the inequality $n! > 5^n$ holds.

Problem 5.7.9. ⋆ Prove the following generalization of the previous problem:

$$\forall m \in \mathbb{N} - \{1\}\,.\ \exists B_m \in \mathbb{N}\,.\ \forall n \in \mathbb{N}\,.\ n \geq B_m \implies n! > m^n$$

Problem 5.7.10. Recall that the **Fibonacci Numbers** are defined by setting $f_0 = 0$ and $f_1 = 1$ and then, for every $n \geq 2$, setting $f_n = f_{n-1} + f_{n-2}$.

Prove the following hold for this sequence:

(a) $\forall n \in \mathbb{N} \cup \{0\}\,.\ f_n < 2^n$

(b) $\forall n \in \mathbb{N}\,.\ f_{n-1}f_{n+1} = f_n^2 + (-1)^n$

(c) $\forall n \in \mathbb{N}\,.\ 1 \leq \frac{f_{n+1}}{f_n} \leq 2$

(d) $\forall n \in \mathbb{N}\,.\ \sum_{k=1}^{n} f_{2k} = f_{2k+1} - 1$

(e) $\forall n \in \mathbb{N}\,.\ \sum_{k=1}^{n} f_k^2 = f_n f_{n+1}$

Problem 5.7.11. In Problem 5.7.1, you proved a formula for the sum of the first n perfect cubes. Specifically, you proved that it is the *square* of the sum of those base numbers.

In this problem, we want you to prove the *converse* of this assertion, that the *only* sequence of numbers with this property is $\langle 1, 2, \ldots, n\rangle$. We will restate this claim below for you to consider, and then prove.

Claim: Suppose $\langle a_i\rangle$ is a sequence of real numbers, i.e. $\forall i \in \mathbb{N}\,.\ a_i \in \mathbb{R}$. Suppose this sequence has the property that

$$\forall n \in \mathbb{N}\,.\ \sum_{k=1}^{n} a_k^3 = \left(\sum_{k=1}^{n} a_k\right)^2$$

Prove that, necessarily, $\forall n \in \mathbb{N}\,.\ a_n = n$, by induction on n.

Problem 5.7.12. (a) Prove that $\forall n \in \mathbb{N}\,.\ 7^n + 7 < 7^{n+1}$.

(b) Prove that $\forall n \in \mathbb{N}\,.\ 3^n + 3 < 3^{n+1}$.

(c) Identify the set S of real numbers r that satisfy $\forall n \in \mathbb{N}\,.\ r^n + r < r^{n+1}$. Prove your claim by induction.

Problem 5.7.13. Prove that for every $n \in \mathbb{N}$, $2^{3^n} + 1$ is a multiple of 3^{n+1}.

Problem 5.7.14. ⋆ Assume $x + \frac{1}{x}$ is an integer. Prove $x^n + \frac{1}{x^n}$ is an integer for all $n \in \mathbb{Z}$

(**Note:** Do this for every $n \in \mathbb{Z}$, not just $n \in \mathbb{N}$!).

Problem 5.7.15. Prove that $n^3 + 5n$ is a multiple of 6 for every $n \in \mathbb{N}$, by induction on n.

Problem 5.7.16. Prove that the following summation equality holds for every $n \in \mathbb{N}$:

$$\sum_{k=n}^{2n-1} 2k + 1 = 3n^2$$

Problem 5.7.17. For every $n \in \mathbb{N} \cup \{0\}$, define

$$s_n = \left(3 + \sqrt{5}\right)^n + \left(3 - \sqrt{5}\right)^n$$

Prove that every such s_n is actually an integer. Also, prove that, in fact, s_n is a multiple of 2^n. (!)

Problem 5.7.18. ► In this problem, we will prove that the familiar **Harmonic Series**, given by

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

is **divergent**; that is, we will prove that the sum of all the terms does not approach a finite limit.

We claim that the following inequality holds for every natural number n:

$$\sum_{k\,=\,1}^{2^n} \frac{1}{k} > \frac{n+1}{2}$$

Call this inequality ★.

(a) Prove that ★ holds true when $n = 1$.

(b) Suppose that $m \in \mathbb{N}$ is arbitrary and fixed, and suppose that ★ holds true for the value $n = m$.

Deduce that ★ also holds for the value $n = m + 1$. Be sure to cite where you use the above assumption about the case where $n = m$.

(c) Think about what you have accomplished so far. Explain how the Harmonic Series cannot converge to any finite limit.

Problem 5.7.19. Prove that the following inequality holds for every $n \in \mathbb{N}$:

$$\sum_{k=1}^{n} \frac{1}{\sqrt{k}} = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots \frac{1}{\sqrt{n}} \geq \sqrt{n}$$

Use this to deduce that the infinite series $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ does **not** converge to a finite limit.

Problem 5.7.20. Prove that the following inequality holds for every $n \in \mathbb{N}$:

$$\prod_{i=1}^{n} \left(i + \frac{1}{i^2}\right) = \left(1 + \frac{1}{1^2}\right)\left(1 + \frac{1}{2^2}\right) \cdots \left(1 + \frac{1}{n^2}\right) < 4 - \frac{1}{n}$$

Problem 5.7.21. In this problem, you will prove the **Well-Ordering Principle** of the natural numbers. This is stated in Theorem 5.5.2, but we'll restate it here:

$$\forall S \in \mathcal{P}(\mathbb{N})\,.\ [S \neq \varnothing \implies (\exists \ell \in S\,.\ (\forall x \in S\,.\ \ell \leq x))]$$

That is, every non-empty set of natural numbers has a **least element**.

You will prove this claim by induction on whether or not a given set S contains n as an element.

We will start the proof for you, and then guide you through the rest:

Let $S \subseteq \mathbb{N}$ be arbitrary and fixed. For every $n \in \mathbb{N}$, define $P(n)$ to be the proposition

$$\text{“ } n \in S \implies [\exists \ell \in S \text{ . } (\forall x \in S \text{ . } \ell \leq x)] \text{ ”}$$

(a) Prove that $P(1)$ holds. (Hint: What's the smallest natural number?)

(b) Let $k \in \mathbb{N}$ be arbitrary and fixed. Write down, using logical notation, a hypothesis which asserts that $P(i)$ holds true for every i between 1 and k (inclusive).

(Hint: This should be easy; just write an "and" statement. Think about what it means.)

Next, suppose $k+1 \in S$. Define $T = S - \{k+1\}$. There are three cases:

(c) Consider the case that $T = \varnothing$. Prove that S has a least element.

(d) Consider the case that $T \neq \varnothing$ and $\forall x \in S \text{ . } x \geq k+1$. Prove that S has a least element.

(e) Consider the case that $T \neq \varnothing$ and $\exists x \in S \text{ . } x < k+1$. Prove that S has a least element.

(Hint: *Here* is where you will need to use an assumption from (b), one of the induction hypotheses!)

Since S has a least element in every case, we deduce that $P(k+1)$ holds. By induction, $\forall n \in \mathbb{N} \text{ . } P(n)$.

(f) Let's show why this proof actually worked! Consider an arbitrary $S \subseteq \mathbb{N}$ such that $S \neq \varnothing$. How do we know S has a least element? That is, which **instance** of the claim $P(n)$ is guaranteed to hold?

(Hint: This would fail if $S = \varnothing \ldots$)

(g) [Bonus] Why didn't we just induct on the size of the set S? Why would that not prove **WOP**?

Problem 5.7.22. Let W be the set of **well-formed strings of parentheses**. Any element w of the set W satisfies one of the following conditions:

(i) w is the string "( )"

(ii) $\exists x \in W$ such that w is the string "(x)" (i.e. w is the string consisting of parentheses around the string x)

(iii) $\exists x, y \in W$ such that w is the string "$x\,y$" (i.e. w is the string consisting of the string y appended after the string x)

For example, "( ) ( )" is a well-formed string in W, because it consists of the valid string "( )" appended after itself. However, "(( )" is not a well-formed string in W, because it does not satisfy any of the above conditions.

(As a more complicated example, we will let you figure out why "(( ) (( )))" is a well-formed string.)

Prove the following statements about this system.

(a) Prove that every element $w \in W$ has an **even** number of parentheses, in total.

(**Hint:** Use a Minimal Criminal argument. Suppose w is a string of the *smallest* odd length ...)

(b) For $w \in W$, let $L(w)$ be the number of left parentheses appearing in w, and let $R(w)$ be the number right parentheses.

Prove that $\forall w \in W \,.\, L(w) = R(w)$.

(**Hint:** Induct on the *length* of the string.)

Problem 5.7.23. What is wrong with the following "spoof" that all pens have the same color?

"Spoof": We claim that all pens have the same color. We will prove this by showing that a set of pens of any **size** has only one color represented amongst those pens. We will provide an inductive argument for this claim, by inducting on the size.

Consider a group of pens with size 1. Since there is only 1 pen, it certainly has the same color as itself.

Now, suppose that any set of n pens has only one color represented inside the group.

Take any set of $n+1$ pens. Line them up on a table and number them from 1 to $n+1$, left to right.

Look at the first n of them, i.e. look at pens $1, 2, 3, \ldots, n$. This is a set of n pens so, by assumption, there is only one color represented in the group. (We don't know what color that is yet.)

Then, look at the last n of the pens; i.e. look at pens $2, 3, \ldots, n+1$. This is also a set of n pens so, by assumption, there is only one color represented in this group, too.

Now, pen #2 happens to belong to both of these sets. Thus, whatever color pen #2 is, that is also the color of every pen in *both* groups. Thus, all $n+1$ pens have the same color.

By induction, this shows that any group of pens, of any size, has only one color represented. Looking at the finite collection of pens in the world, then, we should only find one color. "□"

Problem 5.7.24. An n**-gon** is a convex polygon with n sides. For example, a 3-gon is a triangle, a 4-gon is any rectangle, and so on. ("Convex" means that there are no "indents" in the shape or, equivalently, that if you take any two points inside the shape and draw the line segment between them, that segment does not go outside the shape.)

Prove, by induction on n, that there are $\frac{n(n-3)}{2}$ possible diagonals that can be drawn between the vertices of an n-gon. (Do not count the actual boundary *sides* of the shape as diagonals, only *interior* diagonals.)

5.8 Lookahead

With a whole array of proof techniques and logical knowledge at our disposal now, we could try to go just about anywhere in the mathematical universe. We will choose to explore some particular areas, with the goal of talking about **functions**. We have referred to this idea before, but have not discussed it in a mathematical setting. Over the next two chapters, we will formalize this notion.

Part II

Learning Mathematical Topics

Chapter 6

Relations and Modular Arithmetic: Structuring Sets and Proving Facts About The Integers

6.1 Introduction

Now that we've built a strong foundation of mathematical terminology, concepts, and material, you might be wondering what we're going to talk about next! Well, much like we wanted to *rigorize* the concept of mathematical induction—something we intuitively "understood" but couldn't yet develop in a precise mathematical way—we will, in the next two chapters, set on firmer ground a concept that we have used in passing already, and one that you're likely familiar with in an intuitive sense: a **function**.

To accomplish this, we will start by talking about **relations**. This will lead us into the particular area of **equivalence relations**, which allow us to talk about some qualitative properties of sets. In particular, we will use some equivalence relations on the set $\mathbb{Z}$ to state and prove many interesting properties about integers. This will give us occasion to briefly explore the mathematical branch of **Number Theory**. It is a rich, deep, and broad field, and we will really only skim the surface by stating and proving a few helpful theorems and using them to solve some interesting puzzles and problems. Then, we will move right along to the next chapter and get back to our goal of discussing **functions**.

6.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work

will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to ...

- Define a relation and provide many examples.
- Define and understamd various properties that a relation might have, providing examples of relations that do and do not have these properties.
- Consider a defined relation and discover and prove what properties it has.
- Define equivalence relations and equivalence classes and explain why these are particularly interesting and important examples of relations.
- Consider a defined equivalence relation and categorize its equivalence classes.
- Use a particular equivalence relation on the set of integers to state and prove interesting results in number theory.
- Define the concept of multiplicative inverses, understand what this means in the particular context of modular arithmetic, and apply this idea to prove/disprove the existence of solutions to particular equations.
- State and understand various theorems in number theory, and apply them to solve given problems.

6.1.2 Segue from previous chapter

This chapter doesn't quite follow from the previous one directly, in the way that others have. Instead, we are really moving into a new *part* of the book. From now on, we will be taking all of the mathematical knowledge we have developed thus far and applying it to learn about other interesting areas. We needed to work through all of that other material first, to get to this point. From now on, we will be stating complicated claims and applying proof techniques to prove them. We will provide you with definitions and theorems and expect you to use them to prove other claims. In this sense, this chapter follows from *all* of the previous chapters. We will be putting all of that acquired knowledge, terminology, and experience to good use!

6.1.3 Motivation

You have possibly worked with functions in calculus (differentiating and integrating them) or in a high school algebra course (graphing a function or finding its roots) or even in computer science (coding an algorithm or using recursive programming). But try this: *define* what a function is. How would you explain it to your uncle who's never studied mathematics? How would you explain it to a hyper-intelligent alien? How would you attempt to explain it with the level of rigor that we've provided with mathematical induction? It's not so easy, is it?

To develop a rigorous notion of *functions*, we will first talk about *relations*, a way to compare elements of sets. We will look at plenty of examples and their properties. Then, in the next chapter, we will see that a function is just a particular type of relation! While we talk about relations, we will explore their properties and discover that a particular combination of properties yields a special property. Specifically, we will see that *equivalence relations* yield natural *partitions* of sets, and vice-versa. This result will allow us to state and prove several results about the integers.

6.1.4 Goals and Warnings for the Reader

This chapter will continue our foray into abstract ideas and rigorous mathematics, so it is essential (especially if you feel put off by or uncomfortable with this increasing level of abstraction and the language required as part of it) that you don't get swept up and think that none of this information is "important" or "applicable". All of these concepts will continue to appear throughout this book—and all of mathematics, of course!—so keep that in mind if you find yourself losing focus. We recommend jotting down notes to yourself about what you're learning to remind yourself of what you're doing. When you see a theorem and read through it several times and finally understand it, write down a summary of the theorem in the margin or something so you'll have it later. Draw a little picture to help you conceptualize the important components of an example or theorem. When you read a definition, write down a canonical example and a non-example. After reading a proof, jot down an outline of the steps of the argument so you can "chunk" the concepts and remember (and recall) them more effectively. If you *don't* understand a definition or theorem or proof, make a note about your confusion, too! Take it to a fellow student or smart mathy friend or your TA or professor in office hours and see if they can address your confusion. Most of all, just remember that it takes *time* to digest and internalize these types of abstract concepts and arguments, and it's as important as ever to always work through examples to make sure you follow along in a way that makes sense to *you*. If you can understand something well enough to explain it to someone else, then you're in great shape.

6.2 Abstract (Binary) Relations

6.2.1 Definition

Let's jump right in and start talking about relations. We'll give you a definition and then a bunch of examples.

Definition 6.2.1. *Let A, B be sets. A* **relation** *between A and B is a set of* **ordered pairs***, $R \subseteq A \times B$. Given elements $a \in A$ and $b \in B$, we say a and b are* **related** *if and only if $(a, b) \in R$.*

The set A is called the **domain** *and the set B is called the* **codomain***. The set R is called the* **relation set***.*

If $A = B$, we say R is a relation **on** *A.*

It is also fairly common to write $x\,R\,y$ to mean $(x, y) \in R$. When we have defined a relation in this way, we will stick to the notation $(x, y) \in R$ to indicate the underlying **set** structure. Later on, we will sometimes define relations by using some symbol, like $x < y$ or $x \star y$, and so on.

Remark 6.2.2. A relation, as we defined it here, is also sometimes referred to as a *binary relation*. This is because there are two "inputs" of the relation; the set R consists of ordered *pairs*.

We could generalize this idea to *ternary relations*. That is, given sets A, B, C, we could define a set $R \subseteq A \times B \times C$ to be a ternary relation and say a, b, c are related if and only if $(a, b, c) \in R$. We could generalize this further to relations with n "inputs", as well. In this context, though, we will only consider *binary* relations, so we will use the term *relation* to mean *binary relation*.

Remark 6.2.3. A relation R is often defined by identifying a *property* of elements of A and B (phrased as a variable proposition $P(a, b)$) and setting

$$(a, b) \in R \iff P(a, b)$$

Examples

Example 6.2.4. Let $W = \{\text{English words}\}$ and $L = \{\text{English letters}\}$. Define the relation R by setting

$$(w, \ell) \in R \iff w \text{ begins with } \ell$$

Then, (mathematics,m) $\in R$ and (golf,g) $\in R$ because these are valid words and we have identified their starting letters. For some non-examples, notice that (knowledge,n) $\notin R$ and (you,u) $\notin R$. Furthermore, note that (zyzyxyqy,z) $\notin R$ because zyzyxyqy$\notin W$.

It is often the case that $A = B$, so R defines a relation on pairs of elements from one set. The next example considers this situation.

Example 6.2.5. Let $A = B = \mathbb{Z}$ and define a relation R on $\mathbb{Z}$ by setting

$$(x, y) \in R \iff x \text{ and } y \text{ have the same parity}$$

Then $(2, 8) \in R$ and $(-3, 7) \in R$ and $(-99, -99) \in R$, but $(1, 2) \notin R$ and $(0, -3) \notin R$ and $(\pi, 0) \notin R$ (since $\pi \notin \mathbb{Z}$).

Example 6.2.6. Define a relation L on $\mathbb{R}$ by setting

$$(x, y) \in L \iff x < y$$

Then $(-1, \pi) \in L$ and $(0, 100) \in L$ but $(2, 2) \notin L$ and $(\pi, -1) \notin L$.

Notice that these are *ordered* pairs (which we may forget about since $A = B = \mathbb{R}$) so the order of the elements matters. Indeed, knowing that $(x, y) \in L$ doesn't *necessarily* imply that $(y, x) \in L$, in general. In this example, that implication is always False, in fact!

Recall that we sometimes write $x\, L\, y$ to say $(x, y) \in L$, so let us note that we could say -1 $R\, \pi$ but $\pi \not R$ -1 here, and 2 $\not R$ 2.

The Empty Relation

Remark 6.2.7. The examples we have seen thus far are *interesting* relations, in some sense. Given any $x, y \in \mathbb{R}$, we can determine whether $x < y$ or not by just comparing them and deciding whether that property holds. That is, each of the examples we have seen thus far were defined by saying $(a, b) \in R \iff P(a, b)$ is true for some property $P(a, b)$.

A relation doesn't *necessarily* need to be defined this way, though. For instance, we know $\varnothing \subseteq S$ for *any* set S. Thus, given two sets, we can always define the *trivial relation* by using the fact that $\varnothing \subseteq A \times B$; that is, the trivial relation is the one where no elements are related! This is relatively "uninteresting" for sure, but it still satisfies the definition of a relation, so we allow it.

Any Set of Ordered Pairs is a Relation

Remark 6.2.8. Given sets A, B, **any** subset $R \subseteq A \times B$ defines a relation. It might be difficult (or impossible, perhaps) to identify a property that characterizes that relation.

For instance, if $A = \{1, 3, 5\}$ and $B = \{\star, \heartsuit\}$, then we can define a relation between A and B by setting

$$R = \{\, (1, \star), (5, \heartsuit) \,\}$$

Why is 1 related to $\star$? Why is 3 not related to anything? Who knows? It's just a set of ordered pairs! This is, mathematically speaking, totally fine.

The Equality Relation

Example 6.2.9. Another example of a way to define a relation on any set X is to define the equality relation. That is, let $(x, y) \in R \iff x = y$. Notice that this doesn't depend on what X is or what *types* of objects it contains as elements, merely that it is a *set*.

Similarities Between Relations

Example 6.2.10. Let S be the set of students in your class. Define a relation R_1 between S and $\mathbb{N}$ by saying $(s, n) \in R_1$ if person $s \in S$ is n years old. Write out a few elements of this relation set.

Now, define a relation R_2 on S itself by saying $(s, t) \in S$ if persons s and t are the same age (in years). Write out a few elements of this relation set.

How do the relations R_1 and R_2 compare? Do they somehow "encode" the same information about the elements of the set S? Why or why not? Is there a way we can use R_1 to determine R_2? What about the other way around? Think about this carefully and try to write a few sentences that summarize your thoughts about this. We will adress these questions immediately in the next subsection, but take some time now to investigate on your own!

Relations "Encode" Information

The previous example is meant to illustrate the real use of abstract relations and motivate why we even talk about them at all (besides our overarching goal of rigorizing the notion of a *function*). In some sense, a relation is a way of "storing" information about elements of two sets, or one set; it's a way of *comparing* two elements and declaring whether or not they satisfy some *property*. In a more general sense, though, a relation can provide information about how "well" a set's (or sets') elements behave in terms of a specific property.

For instance, notice that in the previous example, the relation R_1 told us a little "more" about the elements of S. Certainly, R_1 tells us who is the same age as anyone else: we can look for two pairs like (s, n) and (t, n), say, with the same second coordinate. But R_1 also tells us *what* that age is: just look at that second coordinate the pairs share. This does *not* work with R_2. Knowing $(s, t) \in R_2$ just tells us students s and t are the same age; it does *not* tell us what that age is! In that sense, R_1 is a "better" relation and provides "more" information.

There are also reasons why R_2 is "better", too, though! For example, look at one of its nice properties: if $(x, y) \in R_2$, it is *necessarily* true that $(y, x) \in R_2$, as well. This is certainly *not* true of R_1 because when $(s, n) \in R_1$, it doesn't even make sense to say whether or not $(n, s) \in R_1$ because the order of the pair does not match the domain and codomain! Does this property now make R_2 a "better" relation? Well, yes and no. It depends on context and what type of information we want to encode and retrieve. In certain situations, maybe you'd want to use R_1, and other times you'd want to use R_2.

But we're getting slightly ahead of ourselves here! We can't yet describe to you what these properties mean and why they may or may not be desirable. On the whole, though, we are curious about these types of properties and when they do (and do not) hold for *all* pairs of elements in a given set. In the next subsection, we will define and explore several common properties of abstract relations. It is not a guarantee (or requirement) that any relation possess one or more of these properties, but these are the ones that have proven to be interesting and useful in the mathematical and real-world contexts in which relations arise. After that, we'll see lots of examples of relations and discuss how to *prove* these properties hold. While doing this, we'll develop some intuitions for how to work with relations and even figure out the kinds of claims we'd be trying to prove in the first place!

6.2.2 Properties of Relations

Let's start right off by defining a few properties. For each of these properties, every relation either does or does not satisfy it. We encourage you to read each property one by one and try to construct a relation that does satisfy the property, and then one that doesn't. This will help you truly understand the underlying principles of the property and how relations work. (Then, try to define some relations that have some combination of the properties!) After these definitions, we will provide some canonical examples that you might have even come up with yourself! But seriously, do try to come up with some on your own and share any interesting ones that you have!

Definitions: Properties of a Relation on a Set

These properties rely on being able to *reverse* the order of a pair. That is, given $(x, y) \in R$, we might wonder about the pair (y, x); however, the relationship between the domain and codomain demands that $(y, x) \in A \times B$, as well. Thus, we will require $A \times B = B \times A$, which only happens when $A = \varnothing$ or $B = \varnothing$ or $A = B$. (Remember we proved this earlier when talking about sets in Chapter 3!) Since $A = \varnothing$ and $B = \varnothing$ are uninteresting cases, we will assume in these properties that $A = B$ (and $A \neq \varnothing$), so we are defining a relation on *one non-empty set* and comparing its elements with each other.

Definition 6.2.11. *Let A be a set and let R be a relation on A, i.e. $R \subseteq A \times A$.*

- *We say R is* **reflexive** *if*

  $$\forall x \in A \bullet (x, x) \in R$$

  That is, every element is related to itself.

- *We say R is* **symmetric** *if*

  $$\forall x, y \in A \bullet (x, y) \in R \implies (y, x) \in R$$

  That is, the order of the comparison doesn't matter.

- *We say R is* **transitive** *if*

$$\forall x, y, z \in A \bullet [(x, y) \in R \wedge (y, z) \in R] \implies (x, z) \in R$$

That is, relationships can transition through a middle-man.

- *We say R is* **anti-symmetric** *if*

$$\forall x, y \in A \bullet [(x, y) \in R \wedge (y, x) \in R] \implies x = y$$

That is, two different elements can be related in at most one way, or not at all. To see why this is the same statement, let's look at the contrapositive of the conditional statement in the line above:

$$\forall x, y \in A \bullet x \neq y \implies [(x, y) \notin R \vee (y, x) \notin R]$$

Note: it is important to point out that *anti-symmetric* is NOT the same as *not symmetric*. Look carefully at the logical order and quantifiers of the properties to make sense of this. For example, the $\leq$ relation on $\mathbb{R}$ is anti-symmetric but not symmetric. Think about why this is the case.

In fact, try to come up with a relation that is both *anti-symmetric* AND *symmetric*. It actually isn't that hard! We've already mentioned one fundamental relation that has this property.

6.2.3 Examples

Again, try to come up with some relations that do and do not satisfy the four properties we just defined. We will present some nice, canonical examples of relations defined on $\mathbb{N}$ below to give you some concrete ideas to keep in mind. Feel free to add to this list as you come up with simple examples, perhaps defined on other sets like $\mathbb{Z}$ and $\mathbb{R}$.

Example 6.2.12. Throughout this example, relations are defined on the set $\mathbb{N}$.

- Define R_1 on $\mathbb{N}$ by

$$(x, y) \in R_1 \iff x \text{ divides } y$$

(i.e. y is divisible by x, or $\exists k \in \mathbb{N}$ such that $y = kx$. This definition is formally restated below; see Definition 6.2.15)

Then R_1 is reflexive, because $x|x$ since $x = 1 \cdot x$.

The divisibility relation is reflexive.

- Define R_2 on $\mathbb{N}$ by

$$(x, y) \in R_2 \iff x \text{ and } y \text{ have the same parity}$$

Then R_2 is symmetric because if x and y have the same parity then certainly y and x have the same parity.

The "has the same parity" relation is symmetric.

- Define R_3 on $\mathbb{N}$ by
$$(x, y) \in R_3 \iff x < y$$
Then R_3 is transitive because if $x < y$ and $y < z$ then $x < y < z$, so $x < z$.
The "$<$" relation is transitive.

- Define R_4 on $\mathbb{N}$ by
$$(x, y) \in R_4 \iff x \leq y$$
Then R_4 is anti-symmetric because if $x \leq y$ and $y \leq x$ then $x \leq y \leq x$ so $x = y$.
The "$\leq$" relation is anti-symmetric.

Example 6.2.13. Remember that a relation is just a set of ordered pairs. We don't *have* to define it in terms of a *property*. Let's see an example phrased this way, and investigate its properties:

Define the relation R on the set $S = \{a, b, c\}$ by

$$R = \{(a, a), (a, c), (b, c), (c, b)\}$$

Notice that this relation is:

- *Not Reflexive*: $(c, c) \notin R$
- *Not Symmetric*: $(a, c) \in R$ but $(c, a) \notin R$
- *Not Transitive*: $(a, c) \in R$ and $(c, b) \in R$ but $(a, b) \notin R$
- *Not Anti-Symmetric*: $(b, c) \in R$ and $(c, b) \in R$ but $b \neq c$

Example 6.2.14. Let's get a little practice using slightly different notation for relations. Remember that we can also write something like $x\,R\,y$ to mean $(x, y) \in R$.

Define the relation $\star$ on the set S of people in your class by saying, for any $x, y \in S$,

$$x \star y \iff x \text{ and } y \text{ were born in the same month}$$

We claim that this relation is reflexive, symmetric, and transitive. Do you see why?

- The relation is *reflexive* because each person is certainly born in the same month as themself (i.e, $x \star x$).
- The relation is *symmetric* because if person x and person y were born in the same month (i.e. $x \star y$), then certainly person y and person x (just a different order!) were born in the same month (i.e. $y \star x$).
- The relation is *transitive* because ... well, you get the idea, right? We're just appealing to the concept of the word "same" over and over!

What about *anti-symmetry* here? It depends! Are there two *different* people in your class that were born in the same month? If so, this relation is *not* anti-symmetric. However, was everybody in your class born in a different month? If so, this relation *is* anti-symmetric, because no one will be related to anybody but themself! Think about this ...

6.2.4 Proving/Disproving Properties of Relations

When we are presented with a set and a relation on that set, we will immediately wonder whether any of these properties hold. By playing around with some particular elements of the set in question, we can try to conjecture whether or not the relation satisfies a property, and then attempt to prove/disprove it. This sometimes amounts to a bit of "guessing and checking" but, ultimately, to *prove* a property holds, we must prove a statement of the form "For all ... it is true that ...". (Look back at our proof techniques from Section 4.9!) Thus, proving a relations property amounts to taking an arbitrary element (or elements) and arguing about how they are related. To *disprove* such a statement, we would prove its logical negation, which is of the form "There exists ... such that ...." (Again, look back at our proof techniques!) Thus, disproving a property amounts to finding a *counterexample*. Let's look at a couple of examples of proving/disproving relation properties. There are several more examples of these styles of proofs in the exercises, as well.

The "Divides" Relation on $\mathbb{Z}$

We will present (or, perhaps, remind you of) one definition first, because it will be the basis of one of our examples. This is a formal definition of what it means for one integer to *divide* another integer.

Definition 6.2.15. *Let $a, b \in \mathbb{Z}$. We say* **x divides y**, *and write* $\mathbf{x \mid y}$, *if and only if*

$$\exists k \in \mathbb{Z} \bullet y = kx$$

Example 6.2.16. Define the relation R on $\mathbb{Z}$ by

$$(x, y) \in R \iff x \mid y$$

Let's determine whether R satisfies any of the four properties of relations, and then prove/disprove all our claims!

In general, depending on the set and relation in question, you might immediately notice whether or not a property holds, via some intuition or just being able to "see" it right away. If so, great! If not (which is far more likely) we recommend starting a "proof" as if a property actually held, and seeing if you can finish. If you do, well then, you just proved the property! If you struggle at some point, maybe that's because the property doesn't hold, and the point in your proof where you're caught up will give you some insight into finding a counterexample. This strategy doesn't *always* work (maybe you're

struggling through a proof because it's actually challenging, say) but it can be quite helpful, so keep it in mind. We will see it in action in this example, too.

One other strategy—an even simpler one, actually—is to just make a statement out loud, or write down in words what the relation and the property in question is. Sometimes, just making yourself *say* something in plain language, instead of reading abstract symbols on the page, will force you brain into realizing something helpful! We'll see this strategy in action here, too.

- Let's see if R is **reflexive**. What would that actually mean? Let's make ourselves say this out loud. Would we expect that: "Any integer is divisible by itself." Of course! Now, let's try to write that down in the symbolic terms required in a proof.

  Proof. We claim R is reflexive. Let $x \in \mathbb{Z}$ be arbitrary and fixed. Then $x \mid x$ since $x = 1 \cdot x$ and $1 \in \mathbb{Z}$. Thus, $(x, x) \in R$. Therefore, R is reflexive. □

  Voilà! Just thinking out loud helped us realize a fact, and that made it easier for us to write down that statement in mathematical language.

- Let's see if R is **symmetric**. This property is defined in terms of an *implication*, a *conditional statement*. So, let's assume we have an arbitrary related pair, $(x, y) \in R$. Would we necessarily believe $(y, x) \in R$, too? Said another way:

  > Assuming x divides y, can we also say that y divides x?

  This actually seems rather unlikely! Knowing $x \mid y$ tells us that $y = kx$ for some $k \in \mathbb{Z}$, but why would that lead us to believe that $x = \frac{1}{k}y$ means $y \mid x$? What if $\frac{1}{k} \notin \mathbb{Z}$?

  You might be tempted at this point to say something like "Well, $\frac{1}{k}$ is only an integer when $k = 1$ or $k = -1$ so that's that." That isn't quite a full explanation! Remember that to disprove a "For all ..." claim, we need to provide an *explicit counterexample* whenever possible. We don't need to characterize *all* of the cases where the property does and does not hold and try to explain things in generality. We just need *an* example to convince someone that the property does not hold. This is much more illustrative and direct than flailing our arms about and pointing out how some counterexample exists out there somewhere. Let's just show our reader one and move on!

  Proof. Consider $2, 6 \in \mathbb{Z}$. Since $6 = 3 \cdot 2$, we have $(2, 6) \in R$.

  However, writing $2 = \ell \cdot 6$ requires $\ell = \frac{1}{3}$, and $\frac{1}{3} \notin \mathbb{Z}$. Thus, $(6, 2) \notin R$.

  This shows that R is *not symmetric*. □

- Let's see if R is **transitive**. In general, transitivity is typically the most difficult property to think about. This is partly due to the fact that it's defined by a conditional statement with *two* hypotheses, and it uses *three* variables.

  In this specific example, we will assume $x \mid y$ and $y \mid z$, and then wonder whether $x \mid z$ necessarily. Try saying that out loud with words and seeing if you believe it or not.

  It seems like it's true, right? Now, try writing down the hypotheses and the conclusion in mathematical language. Can you see how to piece those together? Try writing out your version of this proof before reading on.

  Proof. Let $x, y, z \in \mathbb{Z}$ be arbitrary and fixed. Suppose $(x, y) \in R$ and $(y, z) \in R$.

  This means $x \mid y$ and $y \mid z$, so $\exists k, \ell \in \mathbb{Z}$ such that $y = kx$ and $z = \ell y$. Let such k, ℓ be given.

  Substituting the first equation into the second, we find that

  $$z = \ell y = \ell(kx) = (k\ell)x$$

  Since $k\ell \in \mathbb{Z}$, as well, we have shown that $x \mid z$. Thus, $(x, z) \in R$, necessarily.

  Therefore, R is transitive. □

- Let's see if R is **anti-symmetric**. This property is also defined by a conditional statement with two hypotheses, so we will assume we have an x and a y and that $x \mid y$ and $y \mid x$. Can we conclude that $x = y$? This question hearkens back to proving that R was not symmetric. Remember, we proved that $x \mid y$ does not necessarily imply $y \mid x$ and, actually, if you thought about it for a moment, it's actually very unlikely that $x \mid y$ *and* $y \mid x$ can both be true. How can this be? Think about it carefully and try to come up with your own proof before you read ours.

  Proof. Let $x, y \in \mathbb{Z}$ be arbitrary and fixed. Suppose $(x, y) \in R$ and $(y, x) \in R$.

  This means $x \mid y$ and $y \mid x$, so $\exists k, \ell \in \mathbb{Z}$ such that $y = kx$ and $x = \ell y$. Let such k, ℓ be given.

  Substituting the first equation into the second, we find that $y = kx = k(\ell y) = (k\ell)y$. We now have two cases.

  Case 1: Suppose $y = 0$. Then we cannot divide both sides by y. Instead, we observe that $x = \ell y = \ell \cdot 0 = 0$, and therefore $x = 0$, as well. Thus, $x = y$ in this case.

  Case 2: Suppose $y \neq 0$. Then we can divide both sides by y. This yields $k\ell = 1$. Since $k, \ell \in \mathbb{Z}$ this means either $k = \ell = 1$ or $k = \ell = -1$.

If $k = \ell = 1$, then $x = \ell y = y$. In the other case ... ⍁

Oh, shucks! This doesn't work! Do you see what happened? In "most" of the cases, we did conclude that $x = y$, but there is actually a possibility that $y = -x$. For instance, when $y = 3$ and $x = -3$, notice that $x \mid y$ and $y \mid x$ but $x \neq y$. THIS is the counterexample we need, and trying to finish our "proof" helped us find it. Perhaps you saw this case coming all along; if so, way to go! Let's wrap this up by presenting that counterexample:

Proof. Consider $x = 3$ and $y = -3$, so $x, y \in \mathbb{Z}$. Notice that $x \mid y$ and $y \mid x$ because $3 = (-1)(-3)$ and $-3 = (-1) \cdot 3$, and since $-1 \in \mathbb{Z}$.

However, certainly $x \neq y$. This shows that R is *not* anti-symmetric. □

As a follow-up question, consider what happens when we define this relation on the set $\mathbb{N}$ instead of $\mathbb{Z}$. What changes? Which properties hold now? Are any answers different than with $\mathbb{Z}$? Do think about this. The answers to those questions will motivate our next subsection.

Constructing a Relation with Specific Properties

One more example before we move on. An interesting "game" to play is to take a set and construct a relation R that satisfies some specific subset of the four properties. (Note: there are 16 different ways that the 4 properties can/cannot hold.) We will ask you questions like this in the exercises, so let's work through one example here.

Example 6.2.17. **Goal:** Let S be the set of students in this class. Define a relation R that is (1) not reflexive, (2) not symmetric, (3) transitive, and (4) anti-symmetric.

To ensure that R is not reflexive, we must make sure that no element is related to itself. To ensure R is not symmetric, we must make sure that whenever a pair $(x, y) \in R$, then $(y, x) \notin R$. To ensure R is transitive, we must make sure whenever $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$. And to ensure R is anti-symmetric, we will think of the contrapositive form of the property's definition, which requires that any pair of elements is related in *at most* one way. This last property is perhaps the hardest to think about; it says that for every $x, y \in S$, either x is related to y but not the other way, or else y is related to x and not the other way, or else x and y just aren't related either way. That is, we must not allow any pairs to satisfy *both* $(x, y) \in R$ and $(y, x) \in R$. (Again, reread the definition of anti-symmetric and write down the contrapositive of the conditional statement and think about why this works.)

Now let's try to construct R so that it satisfies these properties. Property (1) means that our definition can't allow anything of the form "or is equal to", and (2) means that the definition must relate any x and y in "only one way". Thus, we might guess that a *comparison* property, something like the "less than" relation on $\mathbb{Z}$, might work. Let's try it out and attempt to prove/disprove the desired properties.

Let us define R on S by

$$x \, R \, y \iff x \text{ is strictly younger than } y \text{ (in years)}$$

Now, let's explore its properties and make sure they are what we wanted them to be. Try to prove/disprove them on your own before reading our solutions! Also, play around with a *different* relation on S (make one up!) and see how its properties are different. Can you come up with another relation that has the exact same properties as this one?

- R is **not reflexive**. This is because any person $x \in S$ has the exact same age as him/herself, and so $x \not R \; x$.
- R is **not symmetric**. This is because if x is strictly younger than y, then y is strictly *older* than x, so $y \not R \; x$
- R is **transitive**. This is because if x is (strictly) younger than y and y is (strictly) younger than z, then certainly x is (strictly) younger than z.
- R is **anti-symmetric**. This is because for any two people $x, y \in S$, one of them must be younger than the other, or else they are the same age; they cannot *both* be strictly younger than each other. (Essentially, we are ensuring anti-symmetry holds by making sure the *hypothesis* of the conditional statement in that property's definition never holds, so the conditional itself is always True.)

Thus, this relation R satisfies all of the desired properties.

You'll notice that we were not *completely* rigorous in these arguments, but there's a good reason for it. Specifically, we didn't produce *explicit* counterexamples for the properties that fail. It would be best if we could identify two students in your class and show how one is younger than the other but not the other way around. But we don't know who's in your class! That's why we left our arguments as "explaining the existence of something without pointing to it explicitly".

We will point out that, in general, a relation of this form—one defined as $(x, y) \in R \iff x$ is "less" than y (however "less" makes sense in that context)—will be non-reflexive, non-symmetric, transitive, and anti-symmetric. In fact, we can even replace "less" with "greater" and this still holds. To see why this is true, think about the "$<$" relation on $\mathbb{N}$, or $\mathbb{Z}$, or $\mathbb{R}$. Think about the "$>$" relation on those sets. Think about the "is younger than" relation on the set of people, or the "is taller than" relation, or the "has more children than" relation. What about the "$\leq$" relation on $\mathbb{Z}$? How is this different than the "$<$" relation? What properties change?

(These types of questions will be explored a little bit further in the next subsection, where we examine a particular type of relation that behaves like these "$\leq$" and "$\geq$" relations. They are called **order relations**, naturally.)

6.2.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How is *(binary) relation* defined, in terms of sets?

(2) Say we have a relation R defined between A and B. What must true about A and B for us to be able to talk about whether R is **reflexive**, say?

(3) When is a relation **reflexive**? Give an example of a set and a reflexive relation on that set.

(4) When is a relation **symmetric**? Give an example of a set and a symmetric relation on that set.

(5) When is a relation **transitive**? Give an example of a set and a transitive relation on that set.

(6) When is a relation **anti-symmetric**? Give an example of a set and an anti-symmetric relation on that set.

(7) What is the difference between *not symmetric* and *anti-symmetric*?

Give an example of a set and a relation on that set which is both symmetric and anti-symmetric.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Consider the set $A = \{1, 2, 3\}$. For each of the following relations, defined on A or $\mathcal{P}(A)$ as specified, decide whether it is (i) reflexive, (ii) symmetric, (iii) transitive, (iv) anti-symmetric.

Not much justification is required, just a Yes or No and a sentence or two.

(a) R_a on A defined by $R_a = \{\ (1,1), (1,2), (2,1), (2,2), (3,3)\ \}$

(b) R_b on A defined by $R_b = \{\ (1,1), (1,2), (2,2), (2,3), (3,3)\ \}$

(c) R_c on $\mathcal{P}(A)$ defined by $\forall S, T \in \mathcal{P}(A)\mathbf{.}\ (S,T) \in R_c \iff S \cap T = \varnothing$

(d) R_d on $\mathcal{P}(A)$ defined by $\forall S, T \in \mathcal{P}(A)\mathbf{.}\ (S,T) \in R_d \iff S \cap T \neq \varnothing$

(e) R_e on $\mathcal{P}(A)$ defined by $\forall S, T \in \mathcal{P}(A)\,.\ (S,T) \in R_e \iff S \subseteq T$

(2) Define the relation $\star$ on $\mathbb{Z}$ by saying

$$\forall x, y \in \mathbb{Z}\,.\ x \star y \iff 3 \mid x - y$$

(a) Prove that $\star$ is reflexive.

(b) Prove that $\star$ is symmetric.

(c) Prove that $\star$ is transitive.

(Remember that "|" means "divides". Be sure to use the formal definition; see Definition 6.2.15.)

(3) Define the relation $\sim$ on $\mathbb{Z}$ by saying

$$\forall x, y \in \mathbb{Z}\,.\ \ x \sim y \iff 3 \mid x + 2y$$

(a) Prove that $\sim$ is reflexive.

(b) Prove that $\sim$ is symmetric.

(c) Prove that $\sim$ is transitive.

(4) Define the relation T on $\mathbb{R}$ by saying, for any $x, y \in \mathbb{R}$,

$$(x,y) \in T \iff \left(\frac{y}{x} \in \mathbb{R} \ \wedge\ \frac{y}{x} \geq 0 \right)$$

(a) Find $x \in \mathbb{R}$ such that $(x,x) \notin T$. Does this mean T is not reflexive? Why or why not?

(b) Find $x, y \in \mathbb{R}$ such that $(x,y) \in T$ and $(y,x) \in T$. Does this mean T is symmetric? Why or why not?

(c) Find $x, y \in \mathbb{R}$ such that $(x,y) \in T$ but $(y,x) \notin T$. Does this mean T is not symmetric? Why or why not?

(d) Determine whether or not T is transitive, and prove your claim.

(5) Define the relation $\leftrightarrow$ on $\mathcal{P}(\mathbb{N})$ by saying, for any $X, Y \subseteq \mathbb{N}$,

$$X \leftrightarrow Y \iff \left(\ X \subseteq Y \quad \vee \quad X \cap Y = \varnothing \ \right)$$

Prove/disprove each of the four standard properties for this relation (i.e. reflexive, symmetric, transitive, anti-symmetric).

(6) What is wrong with the following "proof" that the symmetric and transitive properties imply the reflexive property?

> Let A be a non-empty set. Let R be a relation on A.
>
> Suppose R is symmetric and transitive. We will show R is reflexive.

Let $x \in A$ be arbitrary and fixed. Define the set T to be

$$\{y \in A \mid (x, y) \in R\}$$

Let $y \in T$ be given. Thus, $(x, y) \in R$.

Since R is symmetric, we can deduce $(y, x) \in R$.

Since R is transitive, and $(x, y) \in R$ and $(y, x) \in R$, we deduce that $(x, x) \in R$.

Since x was arbitrary, we have shown that the reflexive property holds.

6.3 [Optional Reading] Order Relations

Let us discuss some relations that behave like "$\leq$" and have similar inherent properties. This is motivated by the fact that these relations are easily definable on the standard sets of numbers we have—$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$—and they also apply to some other, potentially surprising, situations. We will give a definition first and then consider some examples. We will then use those examples to motivate some interesting properties of order relations and then state and prove those facts.

Definition 6.3.1. *Let R be a relation defined on the set A.*

- *If R is reflexive, transitive, and anti-symmetric, then we say R is a* **partial order** *on A.*
- *If R is reflexive, transitive, and anti-symmetric and, in addition, it satisfies*

  $$\forall x, y \in A \bullet (x, y) \in R \vee (y, x) \in R$$

  then we say R is a **total order** *on A. (That is, a total order is a partial order such that every two elements of the set are comparable one way or the other.)*

This definition tells us what partial and total orders are. The next definition just gives us some useful shorthand for referring to partial and total orders on sets.

Definition 6.3.2. *When R is a partial order on A, we say that the pair (A, R) is a* **partially ordered set***, or sometimes just a* **poset***, for short.*

When R is a total order on A, we say that the pair (A, R) is a **totally ordered set***, or sometimes just a* **toset***, for short.*

We will attempt to explain the reasons for these terms by giving several related examples.

Example 6.3.3. Define the following four relations on $\mathbb{R}$:

$$(x, y) \in R_1 \iff x \leq y$$
$$(x, y) \in R_2 \iff x < y$$
$$(x, y) \in R_3 \iff x = y$$
$$(x, y) \in R_4 \iff \lfloor x \rfloor = \lfloor y \rfloor$$

(Recall that $\lfloor x \rfloor = \max\{a \in \mathbb{Z} \mid a \leq x\}$ is the "floor" of a real number; it is the integer we get by rounding down.)

Which of these are partial orders? Total orders? Neither? Think about this for a few minutes and try to sketch some proofs of your claims, or explain them out loud to a friend/classmate.

Now, here are our thoughts. The relations R_1 and R_3 are both partial orders, but only R_1 is a total order. The relations R_2 and R_4 are neither partial nor total orders (because R_2 is not reflexive and R_4 is not anti-symmetric). The idea behind any type of order relation—partial or total—is that we can somehow *compare* the elements of the set A and assign ... well, an *ordering* to them. Heuristically speaking, a partial order induces "chains" of elements in A so that, along any chain, we can arrange the elements in a line, kind of like the number line and how we usually picture $\mathbb{R}$; for a *total* order, there is only one "chain" and it is all of A.

You might object to the idea that R_2 isn't somehow an "order-like" relation, though, and you'd have a fair point. The only fundamental difference between R_2 and R_1 is that we don't allow equality; quite literally, the phrase "or equal to" is built into the definition of "$\leq$", yet that phrase is left out of the definition of "$<$". This results in R_2 being non-reflexive, but that's it. You might also notice that the relation R_4 doesn't have this same type of relationship with R_1; it seems to be something different (and we'll get to that soon enough). This motivates the following few definitions, wherein a partial or total order can be "relaxed" to a related ordering.

Definition 6.3.4. *Let A be a set and let R be a relation on A. We say R is* **irreflexive** *if and only if*

$$\forall x \in A \bullet (x, x) \notin R$$

Notice that this is *not* the same as merely being *not reflexive*. Think about the quantifiers: reflexivity means *every* element is related to itself, so the logical negation of that means there *exists* at least one element that is not related to itself. Irreflexivity means that *every* element is *not* related to itself.

Definition 6.3.5. *Let A be a set and let R be a relation on A. We say R is a* **strict partial order** *if it is irreflexive, transitive, and anti-symmetric.*

We say R is a **strict total order** *if it is irreflexive, transitive, anti-symmetric, and satisfies the following property:*

$$\forall x, y \in A \bullet x \neq y \implies [(x, y) \in R \vee (y, x) \in R]$$

You might wonder what the connection is to non-strict order relations here. Well, there is a natural way to convert any order relation into a strict one, and vice-versa. We can always define one from the other by either building in, or removing, whether or not elements are related to themselves. The following lemma summarizes how to do this and, in so doing, shows that there are just as many strict orders as there are non-strict ones.

Lemma 6.3.6. *Let (A, R_1) be a partially ordered set. Then the relation S_1 defined by*

$$(x, y) \in S_1 \iff [(x, y) \in R_1 \wedge x \neq y]$$

is a strict partial order on A.

Let (A, R_2) be a totally ordered set. Then the relation S_2 defined by

$$(x, y) \in S_2 \iff [(x, y) \in R_2 \wedge x \neq y]$$

is a strict total order on A.

Let (A, S_3) be a strictly partially ordered set. Then the relation R_3 defined by

$$(x, y) \in R_3 \iff [(x, y) \in S_3 \vee x = y]$$

is a (non-strict) partial order.

Let (A, S_4) be a strictly totally ordered set. Then the relation R_4 defined by

$$(x, y) \in R_4 \iff [(x, y) \in S_4 \vee x = y]$$

is a (non-strict) total order.

Thinking back to the relations R_1 and R_2 defined above on $\mathbb{R}$, it might seem a little odd to define "less than" by "less than or equal to and not equal to". It's certainly wordier! However, this is just a consequence of how we describe "$\leq$" linguistically. Mathematically speaking, it is more natural to speak of reflexive relations and partial and total orders, and then alter those to become strict orders. We will see soon enough—when we talk about *minimal* elements—why reflexivity is a nice property, and this is a reasonable justification for why we would start with a partial order and then amend the definition to allow strict orders, as opposed to the other way around. For now, just notice that R_2 is the strict total order that corresponds to the total order R_1.

Question: is there a strict partial order that corresponds to the partial order R_3? If so, what is it? If not, why?

The relation R_4 is not any type of order relation, strict or otherwise. However, notice that R_4 does nicely "package" the elements of $\mathbb{R}$ together. Essentially, every real number y satisfying $1 \leq y < 2$ is "the same" under this relation. Likewise for every y satisfying $2 \leq y < 3$, and every y satisfying, say, $-5 \leq y < 4$, and so on. Once that "packaging" is accomplished, we "know" that there is an ordering we can assign to those "packages", but no information about that order is encoded in the relation R_4, itself. We would have to do some extra work to impose that ordering. This is why R_4 is not an order relation

of any kind, the way it is defined. However, it is what we call an "*equivalence relation*" because of this nice "packaging" property that partitions the elements of the set into separate classes. This is a notion we will explore in the next section. Once we have established those "packages", we can compare them and order them.

Let's explore some examples in a context other than $\mathbb{R}$. One of these following relations is actually a standard example of a partial order.

Example 6.3.7. Let $S = [3]$ and consider the power set, $\mathcal{P}(S)$. (Remember that the power set of S is the set of all subsets of S.) Define the following relations on $\mathcal{P}(S)$, where $X, Y \subseteq S$:

$$(X, Y) \in R_1 \iff X \subseteq Y$$
$$(X, Y) \in R_2 \iff X \subset Y$$
$$(X, Y) \in R_3 \iff X \cap Y = \varnothing$$
$$(X, Y) \in R_4 \iff X \Delta Y = S$$

Recall that $X\Delta Y$ is the *symmetric difference* of X and Y, and is defined as $X\Delta Y = (X - Y) \cup (Y - X) = (X \cup Y) - (X \cap Y)$.

We claim that R_1 is a partial order but not a total order. Before we go on to prove this claim, consider this challenge problem: Can you define a total order on $\mathcal{P}(S)$? Can you do it in a way that would generalize to the case where $S = [n]$, for some arbitrary $n \in \mathbb{N}$.

Now, to prove that R_1 is a partial order, we must show that it is reflexive, transitive, and anti-symmetric. To also show it is not a total order, we must show that it fails the additional property that says any two elements are somehow comparable. We will accomplish some of these steps, and leave the rest as exercises.

- Let's prove R_1 is anti-symmetric: Let $X, Y \in \mathcal{P}(S)$ and assume $(X, Y) \in R_1$ and $(Y, X) \in R_1$. This means $X \subseteq Y$ and $Y \subseteq X$, and therefore $X = Y$, by standard properties of sets.

- Let's show R_1 is not a *total* order. Consider $X = \{1\} \subseteq S$ and $Y = \{2, 3\} \subseteq S$. Notice that $X \nsubseteq Y$ and $Y \nsubseteq X$, so $(X, Y) \notin R_1$ and $(Y, X) \notin R_1$. That is, X and Y are *incomparable* under this relation.

This relation separates the entire set $\mathcal{P}(S)$ into *chains* that are ordered within themselves, but separate chains may contain elements that are incomparable. For instance, consider the following subsets of $\mathcal{P}(S)$:

$$A_1 = \left\{\varnothing, \{1\}, \{1, 2\}, \{1, 2, 3\}\right\}$$
$$A_2 = \left\{\varnothing, \{1\}, \{1, 3\}, \{1, 2, 3\}\right\}$$
$$A_3 = \left\{\varnothing, \{2\}, \{1, 2\}, \{1, 2, 3\}\right\}$$
$$A_4 = \left\{\{3\}, \{2, 3\}\right\}$$

These sets are not disjoint, so they do not form a *partition* of $\mathcal{P}(S)$. Notice, though, that R_1 does *induce* a *total* order within each subset. By "induce" we mean that we use the same defining property of R_1 but restrict our domain to the set A_1, for instance, instead of all of $\mathcal{P}(S)$. Of course, there are even more sets we could define that are chains under this relation.

Let's formalize this notion and then continue our example

Definition 6.3.8. *Let (A, R) be a partially ordered set, and let $B \subseteq A$. Let $\hat{R}$ denote the relation induced by R on B; that is, we set*

$$\forall x, y \in A \bullet (x, y) \in \hat{R} \iff [x, y \in B \wedge (x, y) \in R]$$

If $(B, \hat{R})$ is a totally ordered set, then we say B is a **chain** *of A (under R).*

With this definition in hand, we see that A_1, A_2, A_3, A_4 are chains of $\mathcal{P}(S)$ under R_1. Now, try proving that R_2 is a (strict) partial order, and then try writing some chains of $\mathcal{P}(S)$ under the relation R_2. How do they compare to chains of $\mathcal{P}(S)$ under R_1?

In the next subsection, we will see why chains are important; specifically, we will look at special properties of partial orders, chains, and total orders, that allow us to find "smallest" and "greatest" elements of subsets.

Before we move on, though, let's see two more related examples of partial orders.

Example 6.3.9. Consider the set $\mathbb{R} \times \mathbb{R}$. We define a relation R on $\mathbb{R} \times \mathbb{R}$ by establishing when a *pair* of real numbers is related to another *pair* of real numbers. Specifically, let's say

$$\big((u, v), (x, y)\big) \in R \iff [u \leq x \wedge v \leq y]$$

One can prove that R is a partial order on $\mathbb{R} \times \mathbb{R}$. We will prove the transitivity property and leave the rest as exercises:

Proof. Let $(u, v), (x, y), (z, w) \in \mathbb{R} \times \mathbb{R}$. Suppose that $((u, v), (x, y)) \in R$ and that $((x, y), (z, w)) \in R$. This means $u \leq x$ and $x \leq z$, so $u \leq z$; also, this means $v \leq y$ and $y \leq w$, so $v \leq w$. Thus, $((u, v), (z, w)) \in R$. This shows R is transitive. □

Hint: to prove R is not a *total* order, we must find a counterexample. That is, we need a pair $(x, y), (u, v)$ such that neither $((x, y), (u, v)) \in R$ nor $((u, v), (x, y)) \in R$. Think about the relation R visually, i.e. geometrically, to come up with such an example.

Think about what the chains are under this relation. Try to describe them geometrically and draw a few representatives.

Example 6.3.10. Let A be the standard English alphabet of 26 letters, and let W be the set of all *finite* strings of letters from A. That is, W is the set of all possible "words", where we allow any combination of letters to be included in our "dictionary". Let's try to define L, the standard *lexicographic* ordering on

W. It helps to represent A as the set [26], where a = 1 and b = 2 and so on, until z = 26. Then, we say a word $w \in W$ is represented by

$$w = (w_1, w_2, \ldots, w_n) \quad \text{where } n \in \mathbb{N} \text{ and } \forall i \in [n]\,.\; w_i \in A$$

Notice that for any two words $v, w \in W$, we can compare them "letter by letter" reading left to right until we reach a difference between them. Wherever that difference occurs, we sort the two words according to the comparison of those two letters. If one word is longer than the other and they have the same letters otherwise, we want to sort the longer one *after* the shorter one, just like "there" comes before "therefore" in the dictionary.

$$(v, w) \in L \iff \text{at the smallest index } i \text{ where } v_i \neq w_i, \text{ we have } v_i < w_i$$
$$\text{(and where a blank space is treated as 27)}$$

Think about why this corresponds to the usual ordering of words in the dictionary. (Could you define this using more rigorous mathematical notation? Try it!)

Now that we've looked at several examples of order relations, we recommend you try several of the exercises to practice identifying these relations and proving their properties. After that, we can move on to talk about many other interesting and useful properties of order relations!

6.3.1 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between a partial order and a total order?

(2) Give an example of a partial order that is not total. Give an example of a total order.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Let $S = [2]$, and define R on $\mathcal{P}(S)$ by $(x, y) \in R \iff x$ has at *least* as many elements as y. Prove that S is not a partial order.

(2) Let $S = [3]$, $T = [2]$, and define $R \subseteq S \times T$ by $(x, y) \in R \iff x \supseteq y$. Prove/disprove each of the four standard properties of relations for R (i.e. reflexive, symmetric, transitive, anti-symmetric.) Use your results to determine whether R is any kind of order relation(s).

6.4 Equivalence Relations

6.4.1 Definition and Examples

Let's shift gears only slightly and talk about another type of relation that satisfies a different subset of the four standard properties of relations. In fact, let's return to a particular relation we mentioned in a previous example: on the set $\mathbb{R}$, define R to be the relation where

$$(x, y) \in R \iff \lfloor x \rfloor = \lfloor y \rfloor$$

(In case you missed it in the Optional Reading , this is seen in Example 6.3.3.)

Notice that this relation is

- *reflexive* because $\forall x \in \mathbb{R}.\ \lfloor x \rfloor = \lfloor x \rfloor$
- *symmetric* because $\forall x, y \in \mathbb{R}.\ \lfloor x \rfloor = \lfloor y \rfloor \implies \lfloor y \rfloor = \lfloor x \rfloor$
- *transitive* because $\forall x, y, z \in \mathbb{R}.\ \big(\lfloor x \rfloor = \lfloor y \rfloor \wedge \lfloor y \rfloor = \lfloor z \rfloor\big) \implies \lfloor x \rfloor = \lfloor z \rfloor$

This particular set of properties has some interesting and useful consequences, so we assign a name to any relation that has these three properties.

Definition 6.4.1. *Let A be a set and R a relation on A. If R is reflexive, symmetric, and transitive, then we say R is an* **equivalence relation**.

That's it! Given any relation R on a set S, all we have to do is go through and prove/disprove these three properties to determine whether R is, in fact, an equivalence relation. Let's run through some of the examples of relations that we have seen already and determine whether they are *equivalence* relations or not, based on what we've proven about them.

Example 6.4.2. (1) Look back at the equality relation on an arbitrary set X that we defined in Example 6.2.9. This is an equivalence relation. Certainly, $(x, x) \in R$, since $x = x$. However, the hypothesis $x\,R\,y$ is false for any $x \neq y$, which makes the conditional statement true. Thus, the only "relevant case" in the symmetric property is the one where $x = y$, in which case, yes, $y = x$, too. Similarly, for the transitive property, if either $x \neq y$ or $y \neq z$, the hypothesis of the defining conditional statement is false, so the statement itself is true; and when $x = y$ and $y = z$, yes, certainly $x = z$. This may not seem like a particularly enlightening development, but it *is* nice to know that there is always at least one equivalence relation we can define on *any* set.

(2) The "divides" relation on $\mathbb{Z}$ is **not** an equivalence relation because it is not symmetric.

(See Example 6.2.16.)

(3) The "is strictly younger than relation" on a (nonempty) set of people is **not** an equivalence relation because it is not reflexive.

(See Example 6.2.17.)

(4) The relation R defined on $\mathbb{Z}$ by

$$\forall x, y \in \mathbb{Z}\text{.}\ x \star y \iff 3 \mid x - y$$

is an equivalence relation because it is reflexive and symmetric and transitive.

(See Exercise 2 in Section 6.2.5.)

This particular example of an equivalence relation will be generalized and discussed in great detail later in this chapter. You might even already recognize it as the "equivalent modulo 3" relation!

Many exercises in this chapter will be of the form "Identify whether or not this definition yields an equivalence relation." These are like problems we have seen before of the form "Prove/disprove the following claim." We need to figure out (somehow) whether we think a given relation is actually an equivalence relation or not; if so, we need to prove that, and if *not* we need to identify which property fails and produce a counterexample to show that. Let's see one example of this to illustrate the idea.

Example 6.4.3. Let $S = \mathbb{N} - \{1\}$ and define $(x, y) \in R \iff x$ and y have a common factor (that is *not* 1, i.e. strictly greater than 1). Let's determine if this relation is, in fact, an equivalence relation by *trying* to prove that it is one, and seeing if the argument breaks down anywhere. If it doesn't, then we have succeeded, and if it does, then we can use that knowledge to construct a counterexample.

First, we notice that $(x, x) \in R$ because x and x have a common factor, x. Also, by the definition of S we have $x > 1$, so R is reflexive. Second, we notice that if $(x, y) \in R$, then x and y have a common factor, some $k > 1$. Then, certainly, switching the letters in the pair doesn't alter this fact: y and x have a common factor $k > 1$, so $(y, x) \in R$, as well, and R is symmetric.

Third, let's assume $(x, y) \in R$ and $(y, z) \in R$. This means x and y have a common factor, call it $k > 1$, and y and z have a common factor $\ell > 1$. Can we use this to find a common factor of x and z? Not necessarily, it seems ... There is no way we can determine that k and ℓ have a common factor, for instance. What if $k = 2$ and $\ell = 3$? Can we identify some values of x, y, z that achieve these common factors, and then verify that x and z have no common factor? Sure, let's consider $x = 2$ and $y = 6$ and $z = 9$. Then $(2, 6) \in R$ and $(6, 9) \in R$ but $(2, 9) \notin R$. This is a counterexample that shows that the transitive property is False for this relation, so it is not an equivalence relation.

We do recommend this method for identifying whether or not a given relation is an equivalence or order relation. Just go through each of the relevant properties—reflexivity, symmetry, transitivity, whatever else you might consider—and **try to prove them**. If you succeed, that's it! If you struggle to prove one of them, use your efforts to identify the problem and see why the property fails. Use this to construct a counterexample for that property.

Motivation

Think about that first example again that we mentioned in this section, where $x \mathrel{R} y \iff \lfloor x \rfloor = \lfloor y \rfloor$. Notice that each real number is related to an integer, specifically the integer we get by *rounding down*. For instance, $1.5 \mathrel{R} 1$ and $\pi \mathrel{R} 3$ and $-1.5 \mathrel{R} -2$. Furthermore, any two real numbers that are related to the same integer are related to *each other*. For instance, $3.5 \mathrel{R} 3$ and $\pi \mathrel{R} 3$, and $3.5 \mathrel{R} \pi$. Because of these observations, we claim that we can "package" all of the real numbers that satisfy $0 \leq x < 1$ into one "cluster" and represent them by a single element of that cluster, say 0. Likewise, we can take all of the real numbers that satisfy $1 \leq x < 2$ and package them into one "cluster" represented by 1. And so on. We didn't *have* to choose 0 and 1 as the representative elements. We could very well have chosen $\frac{1}{2}$ and $\frac{3}{2}$ instead, for example. But the point is that those "clusters" of real numbers are all *related to each other* within the same cluster, and we can represent each of those clusters by one **representative** element.

This observation will lead us directly into the next section, where we will discuss how to formally describe these "clusters". These are called **equivalence classes**. We will then investigate many examples and ascertain some general properties.

Before doing that, we strongly recommend that you play around with some of the examples we have seen already, looking for these kinds of "clusters" and "representatives". For instance, take the relation R defined on $\mathbb{Z}$ by

$$\forall x, y \in \mathbb{Z}\text{.}\ x \star y \iff 3 \mid x - y$$

It **is** an equivalence relation. What are those "clusters" in this case? Can you identify all of them? How many elements are in each cluster? Can you choose representatives?

Try doing the same work with another equivalence relation, like the "was born in the same month" relation on the set of students in your class. (This is an equivalence relation, as you'll realize after a moment's thought.)

A similarly instructive task is to take a **non**-equivalence relation and try to figure out *why/how* it does not have this "cluster" property. For example, take the "divides" relation on $\mathbb{Z}$. Where does it fail to have this property? Does it "come close" at all?

Essentially, do some exploring! It will really help to solidify your understanding of the properties of relations and will make the next section easier to follow.

6.4.2 Equivalence Classes

Definition

Let's say we have an equivalence relation R on a set A. The reason we make the following definition was hinted at in the previous paragraphs. The three properties—reflexivity, symmetry, and transitivity—combine to create a canonical *partition* of a set A. Any elements that are related to each other form a kind of "closed club" or "cluster", and this allows us to refer to any one element of the "club" as a representative, instead of all of them. These "clubs" are called *equivalence classes*, and this idea is explored in the following definition.

Definition 6.4.4. *Let R be an equivalence relation on the set A, and let $x \in A$. The* **equivalence class of** *x (under the relation R) is the set of all elements related to x and is denoted by $[x]_R$. That is,*

$$[x]_R = \{y \in A \mid (x, y) \in R\}$$

Motivation and Examples

The idea behind this definition is that equivalence classes allow us to **partition** the set A into some canonical sets based on the relation R. Look back at Definition 3.6.9 in Chapter 3 to see how we defined a **partition** of a set. (In fact, look at Definition 4.5.11, as well, to see how we restated that definition using logical symbols.) For now, just remember that a partition is a non-empty collection of sets that are pairwise disjoint and whose union is the entire set in question.

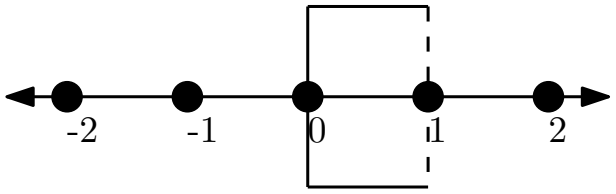
Example 6.4.5. Let's return to the original motivating example from this section. We define the relation R on $\mathbb{R}$ by

$$\forall x, y \in \mathbb{R}\text{.}\ (x, y) \in R \iff \lfloor x \rfloor = \lfloor y \rfloor$$

Let's think about a particular equivalence class, using the definition we just made. Specifically, let's consider

$$\begin{aligned}[0]_R &= \{y \in \mathbb{R} \mid (0, y) \in R\} = \{y \in \mathbb{R} \mid \lfloor 0 \rfloor = \lfloor y \rfloor\} = \{y \in \mathbb{R} \mid \lfloor y \rfloor = 0\} \\ &= \{y \in \mathbb{R} \mid 0 \leq y < 1\}\end{aligned}$$

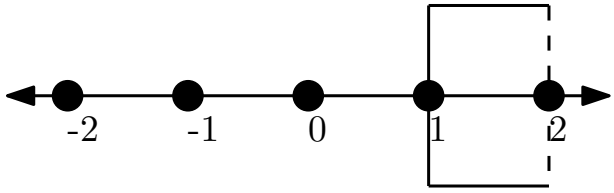
By using the definition of $[0]_R$ from above, the definition of R, and some knowledge of what $\lfloor y \rfloor$ means, we have figured out that the *equivalence class of 0 under the relation R* is this particular interval, between 0 (inclusive) and 1 (exclusive). We can picture that interval like this:



Similarly, we could find that

$$[1]_R = \{y \in \mathbb{R} \mid (1, y) \in R\} = \{y \in \mathbb{R} \mid 1 \leq y < 2\}$$

and picture that set like this:



Notice that these two sets are disjoint (they don't overlap) because the first one does *not* include 1 as an element, but the second one does. Also, notice that *every* real number belongs to *exactly one* interval like this. For example, we can say things like

$$\pi \in [3]_R,\ \mathrm{e} \in [2]_R,\ -1.5 \in [-2]_R,\ \frac{1}{2} \in [0]_R$$

However, notice that the definition of *equivalence class* doesn't say that we have to use exactly one element to *represent* that class. For example, we can say

$$[0]_R = \left[\frac{1}{2}\right]_R$$

because the two sets are equal; they contain the same elements, because any real number whose "floor" is 0 is related to 0 (under R), and therefore also related to $\frac{1}{2}$ under R (because *its* floor is also 0).

Play around with this example some more and try to convince yourself that this partitioning property really works here. In the next part, we will formally *prove* this fact in its full generality, with your help! Because it will be a somewhat abstract discussion, we encourage you to get your hands dirty working with actual examples like this one. Try to define an equivalence relation on another set. What are its equivalence classes? Do you see why they form a partition?

Equivalence Classes Partition the Set

Now that we've explored the idea that equivalence classes *partition* a set, let's formalize this idea. We'll need to make a definition, then we can prove a theorem! The theorem will be, essentially, an "if and only if" style theorem, and we will prove one direction, leaving the other for you as an exercise.

Definition 6.4.6. *Let R be an equivalence relation on the set A. The set of equivalence classes (under R), denoted by A/R, is A* **modulo** *R. That is,*

$$A/R = \{[x]_R \mid x \in A\}$$

Equivalently,

$$A/R = \{X \subseteq A \mid \exists x \in A.\ X = [x]_R\}$$

Let's look at a few examples to get a handle on these ideas before we prove an important result. In each example, let's convince ourselves that we have an equivalence relation, examine the equivalence classes, and think about what the *modulo* operation does.

Example 6.4.7. Consider, again, the relation R defined on $\mathbb{R}$ by $(x, y) \in R \iff \lfloor x \rfloor = \lfloor y \rfloor$. We previously talked about why this is an equivalence relation, so let's examine the equivalence classes.

Any two elements that are related have the same equivalence class, by definition. For instance, $[0]_R = [0.5]_R = [0.999]_R$. Likewise, $[3.5]_R = [3.75]_R$, and $[-\pi]_R = [-4]_R$, but $[\pi]_R \neq [4]_R$. Every real number $x \in \mathbb{R}$ has an associated equivalence class, $[x]_R$, but the idea of the *modulo* operation is to reduce the set $\mathbb{R}$ by considering only as many equivalence classes as are necessary. Since $[0]_R = [0.5]_R = [0.333]_R$ and so on, we can represent all of those identical sets by one set, namely $[0]_R$. Thus, we can say

$$\mathbb{R}/R = \{\dots, [-2]_R, [-1]_R, [0]_R, [1]_R, [2]_R, \dots\}$$

In essence, then, $\mathbb{R}/R$ "is" the set of integers $\mathbb{Z}$. However, we really only feel comfortable writing $\mathbb{R}/R$"="$\mathbb{Z}$ because this equality is not *exact*. In particular, we haven't even rigorously derived the real numbers or the integers yet, only $\mathbb{N}$. Here, we have just observed some kind of "correspondence" between the set of equivalence classes under this relation and the set of integers. We can identify one with the other, and vice-versa, but this doesn't mean they are *equal*, technically speaking.

No matter! The entire point of this example is just to point out that $\mathbb{R}/R$ is a *set* of equivalence classes. Remember that when we write down a set, *order* and *repetition* are irrelavant. That is, $\{1, 3, 5, 3, 1\} = \{1, 3, 5\}$ in the sense of sets. They have the *same elements*, so they are the *same object*. In the current context, we don't need to include *both* $[0]_R$ and $[0.5]_R$ in our set $\mathbb{R}/R$ because they are the same thing; we would be *repeating* that object in our list of elements, and that won't do anything.

In general, what we will be concerned with is identifying what equivalence classes "look like" and coming up with some qualitative description of them. In particular, we will often wonder how *many* equivalence classes there are in A/R. We will also wonder how *big* those classes are. Are they all the same size? Do some only have a couple elements, while others are infinitely large? Why or why not? Do the classes all have roughly the same "description" of their elements?

In this particular example, we have found that all of the equivalence classes in $\mathbb{R}/R$ are quite similar in form. There are infinitely many classes—one for each element of $\mathbb{Z}$—and they are all infinitely big—containing an interval of real numbers. Moreover, all the classes are of the form $[z]_R = \{y \in \mathbb{R} \mid z \leq y < z + 1\}$, for some $z \in \mathbb{Z}$. In that sense, these equivalence classes are all *qualitatively similar*.

Example 6.4.8. Define the relation B on the set of all people S by saying $(x, y) \in B \iff x$ and y were born in the same month. Then, (Leonhard Euler,Henri Poincaré) $\in B$ and (Paul Erdös,Emmy Noether) $\in B$, for example. Why is this

an equivalence relation? Well, any person has the same birth month as themself (reflexivity), and if any two people share a birth month then they ... (duh) share a birth month (symmetry), and if x and y share a birth month and y shares that month with z then x and z share that same birth month (transitivity).

(Note: in general, a relation defined by "has the same ..." or "is the same ..." will be an equivalence relation.)

Equivalence classes under this relation correspond to months! Since we are characterizing people by which month they were born in, an equivalence class is a set of people all born in the same month. For instance, Paul Erdös and Emmy Noether were both born in March, so we can say Emmy Noether$\in$ [Paul Erdös]$_B$. This equivalene class *corresponds* to the month of March, but notice that it is defined in terms of a particular element of the set S (of all people).

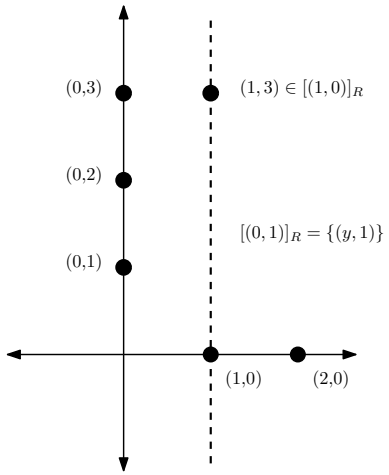
If we define M to be the set of all people ever born in the month of March, then we can say $M =$ [Paul Erdös]$_B$. Taking these observations all together, we can say this: The set of people modulo birth month, written S/B, consists of 12 sets, each corresponding to a different month and containing all of the people born in that month.

Example 6.4.9. Consider the set $\mathbb{R} \times \mathbb{R}$ of all ordered-pairs of real numbers. We define a relation R on $\mathbb{R}\times\mathbb{R}$ by declaring when *two pairs* are related. Specifically, let's say that

$$((x, y), (u, v)) \in R \iff x = u$$

That is, two pairs of points on the plane are related, under R, whenever their first coordinates are the same. Think about why this is an equivalence relation: geometrically speaking, the relation only cares about the vertical line, parallel to the y-axis, that a point lies on. With this in mind, you can easily "see" and explain why R is an equivalence relation, while proving that rigorously just takes a little bit more writing and notation. (Try it!)

This also lets us easily describe and visualize the equivalence classes under this relation. All of the points lying on the same vertical line are packaged together into an equivalence class, and we can index (i.e. keep track of) those classes by just looking at where the line intersects the horizontal axis. That is, for example, $(1, 3) \in [(1, 0)]_R$, because the points $(1, 3)$ and $(1, 0)$ lie on the same vertical line. We can write *every* equivalence class in such a way, $[(x, 0)]_R$, for some $x \in \mathbb{R}$.

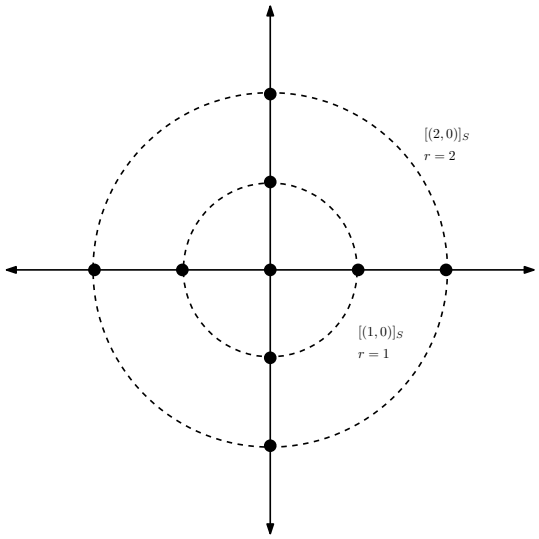


Thus, the set of equivalence classes, $(\mathbb{R}\times\mathbb{R})/R$, is "identical" to the real number line, $\mathbb{R}$, in some sense! We can just collapse all of the points of the plane down to the horizontal axis by ignoring their second coordinates. There is a way of making this idea more precise, mathematically speaking, but we won't be able to truly discuss it formally in this context. Suffice it to say that there's something interesting going on here, in that this relation on $\mathbb{R}\times\mathbb{R}$ yields equivalence classes that are represented by $\mathbb{R}$.

Here's another relation on $\mathbb{R} \times \mathbb{R}$. Define S by setting

$$((x, y), (u, v)) \in S \iff \sqrt{x^2 + y^2} = \sqrt{u^2 + v^2}$$

Remembering some basic geometry and algebra, you might recognize that the expression $\sqrt{x^2 + y^2}$ describes the *distance* from the point (x, y) to the origin $(0, 0)$. (In mathematics, we call such an expression a *metric*.) Thus, the relation says that two points are related whenever they are the same distance from the origin. Visually, this explains why S is an equivalence relation, *and* it shows us that the equivalence classes are circles centered around the origin! Therefore, we can describe the elements of the set $(\mathbb{R} \times \mathbb{R})/S$ by just representing these circles by their one distinguishing feature: their *radius*, some real number $r \geq 0$. Accordingly, the set of equivalence classes, under S, is "identical" to the set of non-negative real numbers!



This is a pretty weird idea, right? We started with a two-dimensional set and related pairs of points and sorted out the equivalence classes and ended up with a one-dimensional set. (Note: We don't have a formal way to define *dimension* here, but we think you have an intuition for what we mean.) Look back at the relation R defined on $\mathbb{R}^2$ above. What if we had defined that relation on just the "right half" of $\mathbb{R}^2$, all of the points whose first coordinate is non-negative. Then, the set of equivalence classes would *also* be "identical" to the set of non-negative real numbers. In what sense would that set be the "same" as $(\mathbb{R} \times \mathbb{R})/S$? Is that even a reasonable question to ask? How might we *prove* statements like this? These are all very interesting questions that we encourage you to think about!

Don't get too distracted by these notions and broad questions, though. The larger point is this: the set of equivalence classes forms a **partition** of the underlying set.

Now that we've seen a few examples, let's state (and prove!) some important results about equivalence relations. Mainly, these theorems present the ideas that we have been alluding to in words all along, namely that an equivalence relation *partitions* a set into its corresponding equivalence classes. Perhaps somewhat surprisingly, though, we have another nice result which says we can do this process in reverse: given any partition, we can define an equivalence relation for it!

Theorem 6.4.10. *Let R be an equivalence relation on the set A. Then the sets belonging to A/R form a partition of A. That is, they are nonempty, they are pairwise disjoint, and their union is A.*

Proof. See Exercise 6.7.13! □

We will guide you through a proof of this result in Exercise 6.7.13 at the end of this chapter. The examples we have examined already should give you

an intuitive understanding of why this theorem is true, but working through the details of the proof will give you a solid understanding of the mathematical rigor behind that.

A Partition Yields an Equivalence Relation

Now, let's move on and look at a similar and important result that is a converse for the previous theorm. To warm up to it, we will first look at one example that will also give us a sketch of the theorem's proof.

Example 6.4.11. Consider the set $S = [6]$. Define the collection of sets

$$\mathcal{F} = \{\, \{1,4\}\, \{2,3,5\}\, , \{6\}\, \}$$

Notice that $\mathcal{F}$ is a partition of S because the sets are disjoint, none are empty, and their union is S. Wouldn't it be nice if there were some equivalence relation R that yielded these sets when we considered S/R? It turns out that there is! Of course, we may not be able to *define* it in a nice way, like the relations we have seen so far, which are usually defined as "$(x, y) \in R \iff x$ and y share some common property". However, having the partition in hand already allows us to define the relation *in terms of the partition*. Specifically, the partition sets *are* the equivalence classes. The partition itself builds in the equivalence class structure, and we can just define an equivalence relation R by saying $(x, y) \in R \iff x$ and y belong to the same partition set.

In this example, we would define $S_1 = \{1, 4\}$ and $S_2 = \{2, 3, 5\}$ and $S_3 = \{6\}$. Then, we define the relation R by

$$(x, y) \in R \iff \exists i \in [3]\,.\, (x \in S_i \wedge y \in S_i)$$

Think about why this works. Do you see why this is an equivalence relation? Do you see what the equivalence classes are?

Now we're ready to state the theorem and prove it.

Theorem 6.4.12. *Let S be a set and let $\mathcal{F}$ be a partition of S. Then there exists an equivalence relation R such that $S/R = \mathcal{F}$.*

As we hinted at above, this result hinges entirely upon the fact that a partition is a collection of sets that are *precisely* the equivalence classes we want to define. All that we need to do is prove that the relation "x and y are related if and only if they belong to the same partition set" is an equivalence relation. It's not too hard! Try to sketch the details of the proof before reading our version!

Proof. Let $\mathcal{F}$ be a partition of S. This means we have an index set I, and

$$\mathcal{F} = \{S_i \mid i \in I\}$$

where the sets S_i satisfy $S_i \subseteq S$ and $S_i \neq \varnothing$ and

$$\bigcup_{i \in I} S_i = S \qquad \text{and} \qquad \forall i, j \in I\,.\, i \neq j \implies S_i \cap S_j = \varnothing$$

Let's define the relation R on S by

$$(x, y) \in R \iff \exists i \in I \,.\, \big(x \in S_i \wedge y \in S_i\big)$$

We will now prove R is an equivalence relation.

- Let $x \in S$ be arbitrary and fixed. Since the sets S_i cover S, we know $\exists i \in I \,.\, x \in S_i$. Let such an i be given.

  Certainly, then $x \in S_i$ and $x \in S_i$, so $(x, x) \in R$. Therefore, R is reflexive.

- Let $x, y \in S$ be arbitrary and fixed. Suppose $(x, y) \in R$. This means $\exists i \in I \,.\, \big(x \in S_i \wedge y \in S_i\big)$. Let such an i be given.

  Certainly, then, we have $y \in S_i \wedge x \in S_i$. Thus, $(y, x) \in R$, as well.

  Therefore, R is symmetric.

- Let $x, y, z \in S$ be arbitrary and fixed. Suppose that $(x, y) \in R$ and $(y, z) \in R$. This means $\exists i \in I \,.\, \big(x \in S_i \wedge y \in S_i\big)$ and $\exists j \in I \,.\, \big(y \in S_j \wedge z \in S_j\big)$. Let such i, j be given.

  Notice that $y \in S_i \wedge y \in S_j$. Since $S_i \cap S_j = \varnothing$ for any *distinct* i, j, it must be that $i = j$. (Otherwise, $y \in \varnothing$, which is impossible!)

  Accordingly $x \in S_i$ and $y \in S_i$ and $z \in S_i$. Thus, $(x, z) \in R$.

  Therefore, R is transitive.

Since all three properties hold, R is an equivalence relation!

The equivalence classes of S modulo R, S/R, are of the form $[x]_R$, where $x \in S$. Since $\mathcal{F}$ is a partition of S, $x \in S_i$ for some i. Thus, $[x]_R = S_i$ for some i. Therefore, all the equivalence classes are equal to some set S_i.

Likewise, any set $S_i \neq \varnothing$, so $\exists x \in S_i$, and thus there is a corresponding equivalence class $S_i = [x]_R$. Therefore, every equivalence class is a set of the form S_i, and vice-versa. □

This shows that any partition corresponds nicely to an equivalence relation, and its classes!

6.4.3 More Examples

Now that we have these two theorems in hand, let's work with some examples of relations. For each, we'll try to figure out whether it is an equivalence relation or not. If it is, we can describe its equivalence classes. If it's not, we can try to invoke one of the theorems and see *why* it isn't.

Example 6.4.13. Let's start with an easy one. Look back at the equality relation that we defined in Example 6.2.9. We explained already that "=" is an equivalence relation on any set. Specifically, it partitions a set into the equivalence classes which are ... well, the elements of the set themselves! That is, on the

set $\mathbb{N}$, say, $[1]_= = \{1\}$ and $[2]_= = \{2\}$, and so on. The equivalence classes are all *singletons* (sets with one element each).

Example 6.4.14. Let's do another fairly easy example. Look back at the parity relation on $\mathbb{Z}$ that we defined in Example 6.2.5. It is an equivalence relation, so let's prove that now.

Proof. Let $a, b, c \in \mathbb{Z}$ be arbitrary.

First, notice that $(a, a) \in R$ because a has the same parity as itself. Thus, R is reflexive.

Second, assume that $(a, b) \in R$, so a and b have the same parity; certainly, then, b and a have the same parity, so $(b, a) \in R$. Thus, R is symmetric.

Third, assume $(a, b) \in R$ and $(b, c) \in R$. If a is odd, we can deduce that b is odd and then c is odd; similarly, if a is even, we can deduce that b is even and then c is even. In either case, a and c have the same parity, so $(a, c) \in R$ necessarily. Thus, R is transitive.

Since R is reflexive, symmetric, and transitive, R is an equivalence relation. □

This means the set of equivalence classes $\mathbb{Z}/R$ forms a partition of $\mathbb{Z}$. Let's identify that partition.

Consider $[0]_R$. This is the set of all integers related to 0, i.e. the set of integers that have the same parity as 0, namely all the *even* integers. Thus, in this case

$$\mathbb{Z}/R = \{O_\mathbb{Z}, E_\mathbb{Z}\}$$

where $O_\mathbb{Z}$ is the set of odd integers and $E_\mathbb{Z}$ is the set of even integers. There are two equivalence classes, each infinitely big.

Example 6.4.15. Look back at the order relation on $\mathbb{R}$ that we defined in Example 6.2.6. Is this an equivalence relation? To figure this out, we can check each property in the definition. Notice that, whatever $x \in \mathbb{R}$ is, we have $(x, x) \notin R$ because $x \not< x$. Thus, R is not reflexive, and therefore not an equivalence relation. (It is also true that R is not symmetric, but it *is* transitive.)

Why does it make sense that this strict order relation would not be an equivalence relation? Why do we want an equivalence relation to be reflexive? Think about the concept of an *equivalence class*; an equivalence relation should place the elements of the whole set into a partition, where we can identify any partition set by an element that belongs to it. With a relation that is not reflexive, we would then have some elements that don't belong to *their own* "equivalence classes", surely an undesirable situation!

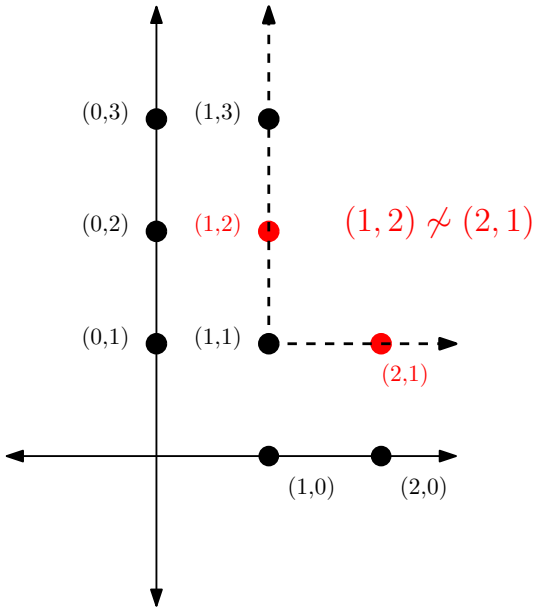
(Follow-up question: What about the order relation $\leq$, which is reflexive; is this an equivalence relation? Why or why not?)

Put another way, we can see that the relation "$<$" on $\mathbb{R}$ does not separate the real numbers into a partition. Because of this, and thinking about the *contrapositive* of Theorem 6.4.10, we conclude that "$<$" *cannot* be an equivalence relation.

Example 6.4.16. Define the relation $\sim$ on $\mathbb{R} \times \mathbb{R}$ by

$$(x, y) \sim (u, v) \iff x \leq u \wedge y \leq v$$

Without even examining its properties, let's see if we can identify whether it is an equivalence relation or not. To do this, let's take a specific element of the set and look at all of the elements related to that specific one. For the picture below, we will use $(1, 1)$ as the specific element.



Notice that the defining condition of $\sim$ asks that a point lie "above and to the right" of another one for the two to be related. Also, notice that the inequalities are "$\leq$" so the second point doesn't have to be *strictly* above or to the right.

Thus, $(1, 2) \sim (1, 1)$ as we can see from the picture (also observing that $1 \leq 1 \wedge 1 \leq 2$). Also, $(1, 1) \sim (2, 1)$, for similar reasons. Accordingly, the points $(1, 2)$ and $(2, 1)$ are both related to $(1, 1)$ and so for this relation $\sim$ to be an *equivalence* relation, we *would* require that $(2, 1)$ and $(1, 2)$ be related *to each other*. This is because they would both have to belong to the "equivalence class" of $(1, 1)$. However, notice that $(1, 2) \not\sim (2, 1)$, unfortunately! The second point lies strictly "below and to the left" of the first one, so it does not satisfy the defining condition of $\sim$.

This means that the set of all elements related to $(1, 1)$ does **not** form a "closed club". Mathematically speaking, this set of elements is not an equivalence class. Therefore, $\sim$ is **not** an equivalence relation.

Now, try to identify which properties $\sim$ does and doesn't have. Is it reflexive? Symmetric? Transitive? Why or why not? In so doing, you will prove again that $\sim$ is not an equivalence relation. Wasn't it helpful to have already

figured out beforehand that it *isn't*? We recommend, in general, doing something similar when you are faced with a defined relation. Can you figure out what the "equivalence classes" might be? If so, then you've developed some intuition for how and why the relation is an equivalence relation, and it will help you describe the equivalence classes. If not, then you've developed some intuition for how to disprove such a claim.

[Optional Reading] How $\mathbb{Z}$ comes from an Equivalence Relation on $\mathbb{N} \times \mathbb{N}$

Remember that crazy exercise from Chapter 3 that had you prove something about a set of pairs of pairs of natural numbers, and we claimed that was proving something about the existence of the integers? What was that all about? Look back at the exercise now, Exercise 3.11.22. You'll see that the last three parts of the problem have you prove that the set R we defined is an **equivalence relation** on the set P. (The underlying set was $P = \mathbb{N} \times \mathbb{N}$.) Look at that! You proved R is reflexive, symmetric, and transitive.

What that exercise showed is that (essentially, we are glossing over some details here) any negative integer is represented as the **equivalence class** of pairs of integers whose difference is that negative integer. That is,

$$-1 \text{ “}=\text{”} [(1,2)]_R = \{(1,2),(2,3),(3,4),\dots\}$$

and, for another example,

$$-3 \text{ “}=\text{”} [(1,4)]_R = \{(1,4),(2,5),(3,6),\dots\}$$

This is only an intuitive explanation and not rigorous, mathematically speaking, but that's the main idea!

6.4.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What properties does an *equivalence relation* have to satisfy?

(2) What is an equivalence class? What must be true about all of the elements in one equivalence class?

(3) Given a set S and an equivalence relation R on S, what must be true about the set of equivalence classes?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Look back at the relation defined in Exercise 2 in Section 6.2.5. There, we define the relation $\star$ on $\mathbb{Z}$ by setting

$$\forall x, y \in \mathbb{Z}\text{.}\ x \star y \iff 3 \mid x - y$$

You proved there that it is, indeed, an equivalence relation.

Now, describe the equivalence classes in $\mathbb{Z}/\star$. How many are there? How "big" are they? Can you list their elements or describe them somehow?

(2) Look back at the relation defined in Exercise 3 in Section 6.2.5. There, we defined the relation $\sim$ on $\mathbb{Z}$ by saying

$$\forall x, y \in \mathbb{Z}\text{.}\ \ x \sim y \iff 3 \mid x + 2y$$

You proved there that it is, indeed, an equivalence relation.

Now, identify and describe the equivalence classes in $\mathbb{Z}/\sim$. How many are there? How "big" are they? Can you list their elements or describe them somehow?

Compare this to the previous exercise. What do you notice?

(3) Consider the set $[5] = \{1, 2, 3, 4, 5\}$. Define the relation $\approx$ on $[5]$ by setting, for any $x, y \in [5]$

$$x \approx y \iff |x^2 - y^2| \leq 5$$

For every $x \in [5]$, let $S(x)$ be the set of all elements $y \in [5]$ such that $x \approx y$.

(a) Write out all the elements of the sets $S(1), S(2), S(3), S(4), S(5)$.

(b) Can you determine whether or not $\approx$ is an equivalence relation by looking at these sets? How?

(c) Prove whether or not $\approx$ is an equivalence relation by proving/disproving the reflexive, symmetric, and/or transitive properties.

(4) Consider the set $\mathbb{N} \times \mathbb{N}$. Define the relation $\sim$ on this set by setting

$$(a, b) \sim (c, d) \iff a + b = c + d$$

Determine whether or not this is an equivalence relation. If it is, describe its equivalence classes visually.

6.5 Modular Arithmetic

A natural and common equivalence relation that you may have already seen and worked with before is that of *congruence*, in the context of the integers. This is a direct generalization of the "even/odd parity" equivalence relation, which classifies integers based on one specific property. Here, we expand this idea by defining a few properties of integers and then defining a bunch of relations. We will also go through some interesting results that become easy to prove (or provable at all!) by using these relations.

6.5.1 Definition and Examples

Divisibility

We'll start with a definition that we have seen a few times already.

Definition 6.5.1. *Let $a, b \in \mathbb{Z}$. We say that a* **divides** *b if b is evenly divisible by a, i.e. $\exists k \in \mathbb{Z}$ such that $b = ak$, or equivalently, $\frac{b}{a} \in \mathbb{Z}$ (except for the case where $a = b = 0$). We denote this by $a \mid b$.*

Notice that this definition says that every integer divides 0 (e.g., $5 \mid 0$) but 0 doesn't divide anything except itself (e.g., $0 \nmid 5$ but $0 \mid 0$). Think about how this makes sense with your intuitive understanding of "divides" and also how it satisfies the given definition. Also, notice that negative numbers are accounted for here, since the existence quantifier takes an *integer* $k \in \mathbb{Z}$. Thus, $-2 \mid 4$ and $8 \mid -24$, as well.

Now, a statement like $2 \nmid 5$ tells us some information about how the integers 2 and 5 are related, but it doesn't say everything. We know that there is *no* possible integer k that satisfies $2k = 5$, but it doesn't say how *close* we can get. Certainly $k = -100$ is a bad estimate, but $2 \cdot 2 = 4$ and $2 \cdot 3 = 6$ are pretty close to 5 ... This seems obvious with small numbers like this where we can check by hand, but what about huge numbers? We know that $7 \nmid 100000$ (Why? Think about primes ...), but how can we approach this "find the k that makes $7k$ the closest possible to 100000" problem? How do we know there's even a specific answer? Might there be two equally "reasonable" answers, like with $2 \nmid 5$?

Regarding the second question, about plurality, we'd like to restrict ourselves so that there is only one reasonable answer. This comes from a desire for simplicity, not having to worry about finding another answer after finding one. Accordingly, we will follow the *The Price Is Right* standard: we want the *closest* answer without *going over*. With the example $2 \nmid 5$, we consider $k = 2$ to be the best estimate since $4 < 5$. Likewise, with the example $7 \nmid 100000$, we consider $k = 14285$ to be the best estimate because $7 \cdot 14285 = 99995$. (Notice that, in this case, there is a "closer" estimate that does go over, but we don't consider it.)

This now motivates how to come up with such estimates. Given $a, b \in \mathbb{Z}$, we can just look at larger and larger multiples of a until we go too far, past b; the multiple right before that will be the best. The "accuracy" of the estimate

has to be some amount between 0 and $a-1$, where 0 happens if $a \mid b$ is actually true. (Note: the idea of "going too far" is the order relation $>$, so think carefully about how this applies to negative integers. For instance, $2 \nmid -3$ and $2 \cdot -2 = -4$ is considered the best estimate, since $-4 \leq -3$.) The following Lemma encapsulates these ideas of iteratively looking at multiples of a until we find the best estimate for b, and declares that there is always a unique solution, under the constraints we have agreed upon.

Lemma 6.5.2 (The Division Algorithm)**.** *Let $a, b \in \mathbb{Z}$. Then $\exists!\, k, r \in \mathbb{Z}$ such that $ak + r = b$ and such that $0 \leq r \leq a-1$. Said another way, given any two integers, there is always a unique multiple of a that is closest to b without being greater, and there is a corresponding unique remainder. We call this r the "remainder of b upon division by a" or "the remainder when b is divided by a".*

It is this concept of *remainder* that we will most frequently use from this result. Specifically, we will compare remainders of two divisions and define a relation based on those remainders. We'll see those details shortly. First, we want you to prove this important Lemma!

Proof. Left for the reader as Exercise 6.7.14. □

The reason this is called the Division *Algorithm* is that there is an implied *process* by which we can actually *find* these multiples and remainders. This method is, naively but effectively enough, *repeated subtraction*. That is, given a and b, we can just keep subtracting a from b—finding $b-a$ and then $b-2a$ and then $b-3a$ and then $\ldots$—until we are left with a remainder between 0 and a.

Example 6.5.3. Let's see this process played out, just to show you the idea. Let's use $a = 8$ and $b = 62$. We continually subtract 8 from 62, finding

$$62,\ 54,\ 46,\ 38,\ 30,\ 22,\ 14,\ 6$$

We stop at 6 because it satisfies $0 \leq 6 < a = 8$. This tells us $r = 6$. We also notice that we subtracted a from b seven times, since there are eight terms in our list, with the first one being just $b - 0 \cdot a$. Thus, we can write

$$\underbrace{62}_{b} = \underbrace{7}_{k} \cdot \underbrace{8}_{a} + \underbrace{6}_{r}$$

The main point here is that there *exists* such a way find this remainder, and that it is unique. With that result in hand, let's use it to define some relations on $\mathbb{Z}$. We will go on to show that these are all equivalence relations and, more specifically, see how useful their equivalence *classes* are!

Congruence modulo n

Definition 6.5.4. *Let $n \in \mathbb{N}$. We define a relation R_n on $\mathbb{Z}$ by saying $(a, b) \in R_n$ if and only if a and b have the same* remainder *upon division by n.*

Equivalently, we say $(a, b) \in R_n \iff n \mid a - b$.

Notationally, we also write this as

$$a \equiv b \mod n$$

and read this as "a and b are **congruent modulo** *n. (Verbally, we usually shorten "modulo" to "mod".)*

Remark 6.5.5. We have stated in our definition that saying "a and b have the same remainder upon division by n" is *equivalent* to saying "$n \mid a - b$". Why is this the case? It's not *by definition*; it requires a little proof. We will ask you to do this later, in Exercise 6.7.15.

Remark 6.5.6. In practice (i.e. in solving problems and proving other results), we will use this definition as follows: knowing that $a \equiv b \mod n$ guarantees that we can express a as a multiple of n plus b.

Let's see why that works. Let's say their shared remainder is r. This means there exist $k, \ell \in \mathbb{Z}$ such that

$$a = kn + r \quad \text{and} \quad b = \ell n + r$$

(They have the same remainder, but they might not have the same *multiple* of n.) Subtracting to solve for r and then setting them equal, we find

$$a - kn = b - \ell n$$

and then adding and factoring tells us

$$a = (k - \ell)n + b$$

Look at that! The term $(k - \ell)n$ is a multiple of n, and the second term is just b itself. This tells us a is a multiple of n plus b.

In general, b might not be the remainder of a upon division by n, itself; in particular, this happens when b does *not* satisfy the requirement $0 \leq r \leq a - 1$ that we ask of remainders.

Let's sum up this remark by writing down the form of this definition that we will invoke in the future. This is the statement we will be refer to when we use the definition of *congruence modulo n* in a proof or an example:

$$a \equiv b \mod n \iff \exists m \in \mathbb{Z}\text{.}\ a = mn + b$$

Example 6.5.7. Let's figure out what these relations "look like" by considering some small values of n.

- Let $n = 1$. What does the relation R_1 look like? This is actually a somewhat silly question because the remainder when any integer is divided by 1 is 0, so every integer is related to every other integer. That is, $\forall x, y \in \mathbb{Z}\text{.}\ (x, y) \in R_1$. Because this is relatively uninteresting, mathematicians would hardly ever speak of "mod 1".

- Let $n = 2$. The relation R_2 is precisely the "parity relation" we defined before. Think about why this is true. When we divide any integer a by 2, the only possible remainders are 0 and 1. If a and b share a remainder of 0 upon division by 2, then they are both even; if they share a remainder of 1, they are both odd. (Think about how this corresponds to our definition way back in Chapter 3, where *odd* and *even* were defined in terms of *existence* claims: e.g. x is even if and only if $\exists k \in \mathbb{Z}$ such that $x = 2k$. This is exactly what the division algorithm result says here: x is even if and only if its remainder upon division by 2 is 0, because we can find an integer such that $x = 2k$.)

  Now, think about the equivalent formulation of congruence. If two integers are even, what can we say about their difference? That's right, it's also even! Here, this means $a \equiv b \bmod 2 \iff a - b \mid 2$; i.e. a and b are both even (or both odd) if and only if their difference is also even. (Note: we haven't yet *proven* that this other formulation is truly equivalent to the definition in terms of remainders. We will do that immediately after this example.)

- Let $n = 3$. Then, for example, $0 \equiv 9 \bmod 3$ and $-1 \equiv 2 \bmod 3$ and $4 \equiv 28 \bmod 3$. In general, we can also string together several statements of congruence, as long as we tack on "mod 3" (or what have you) at the end of the line. When we do this, it is understood that the entire preceding line is considered only modulo 3. For instance, the following line is valid, notationally, and true, mathematically:

  $$-100 \equiv -1 \equiv 8 \equiv 311 \equiv -289 \equiv 41 \mod 3$$

  (We're not sure why you'd have to write such a statement, but we're just pointing out that it's perfectly okay to do so!)

- Let $n = 10$. The remainder of a natural number divided by 10 is just its last digit, its *ones* digit! This helps us compare two numbers modulo 10 easily. For instance, $12 \equiv 32 \equiv 448237402 \bmod 10$ but $37457 \not\equiv 38201 \bmod 10$.

  This is different when we consider *negative* numbers, though. The reason is that we defined remainders by taking the largest multiple *without going over*. Then, for example, $-1 \equiv 9 \bmod 10$; this is because $-1 = (-1)\cdot 10 + 9$ and $9 = (0) \cdot 10 + 9$. They share a remainder of 9 that needs to be *added* on to some multiple of 10. Think about the details behind the following True statements:

  $$-3 \equiv 17 \equiv -33 \equiv 107 \mod 10$$

Notation

One important comment about *notation*: in mathematics, **mod** is a relation, not an operator or function. In computer science and programming, you might see something like "5 mod 3 = 2" to say that "the remainder when we divide

5 by 3 is 2". (In many languages, this might be expressed as `5%3==2`.) You won't see us write anything like that here. Rather, we use mod to indicate some kind of **equivalence**, using $\equiv$ along the way because the numbers we are talking about are not necessarily *equal*. If we express a chain of equivalences that makes sense modulo some natural number n, we will write "mod n" at the end of the line to indicate that. In this sense, mod is more like a *modifier* that we write to say "All of the statements made on this line are only meant to be considered in the sense of remainders when dividing by n". Thus, we can write something like

$$100 \equiv 97 \equiv 16 \equiv 4 \equiv z \cdot w \equiv 1 \equiv x - y \equiv -2 \equiv -8 \mod 3$$

which says all of those numbers and expressions are equivalent when considered modulo 3. We aren't asserting that they are equal, nor that they are necessarily equivalent modulo anything else. The "mod 3" at the end says, "We are working inside the universe of the integers modulo 3, and nowhere else."

(Question: Can you find $x, y, z, w \in \mathbb{Z}$ that make the line above `True`?)

Three Important Lemmas

Here, we will ask you to prove two important results; namely, you will prove that congruence modulo n can be thought of equivalently in terms of *divisibility*, and that these relations are *equivalence relations*. Work through these corresponding exercises *now*, while you are reading this section. The following section—where we talk about the equivalence *classes* under these relations—will make much more sense to you if you have already worked through these details. After these two proofs, we will present one more result and prove it for you. The last example we see before talking about equivalence classes will be an interesting arithmetic problem that is easily solvable using congruences, but not exactly easy if you want to do it "by hand", so to speak.

Lemma 6.5.8. *The two formulations of congruence modulo n given in Definition 6.5.4 are indeed equivalent. That is, for every $a, b \in \mathbb{Z}$ and for every $n \in \mathbb{N}$,*

$$a \text{ and } b \text{ have the same remainder upon division by } n \iff n \mid a - b.$$

Proof. See Exercise 6.7.15. □

Lemma 6.5.9. *For any $n \in \mathbb{N}$, R_n is an equivalence relation on $\mathbb{Z}$.*

Proof. See Exercise 6.7.16 □

Thank you for proving those Lemmas! ☺We now know that congruence modulo n is an equivalence relation (so we *can* talk about equivalence classes) and that we can always figure out whether two integers (say a and b) are congruent modulo n by just determining whether $a - b$ is a *multiple* of n. This will be a convenient way of reading off proposed congruences and assessing whether or not they hold.

The next lemma tells us that we can perform **arithmetic**—addition and multiplication—in the context of "mod n" and rest assured that results still work correctly. What if we had two *equations*, two statements of equality about integers, and we added them together? We know the resulting equality still works, right? That is, if we know $a + b = c$ and $d + e = f$, then we can add them and know that $a + b + d + e = c + f$. This lemma says the same thing works with congruence modulo n instead of equations. Likewise, we can *multiply* congruences and rest assured they preserve congruence.

The proof of this lemma is not too difficult, but we will prove it for you, sincer we've been having you do a lot of the work lately.

Lemma 6.5.10 (Modular Arithmetic Lemma, or MAL)**.** *Let $n \in \mathbb{N}$ be given. Let $a, b, r, s \in \mathbb{Z}$ be arbitrary and fixed. Suppose that $a \equiv r \bmod n$ and $b \equiv s \bmod n$. Then*

$$a + b \equiv r + s \mod n$$
$$a \cdot b \equiv r \cdot s \mod n$$

(If you think about it, this Lemma tells us we can just work with remainders. Whatever a, b we are given, we can just reduce them to their remainders, r and s, and work with those instead. The idea is that $0 \leq r, s \leq n - 1$, so they are guaranteed to be *small*, compared to a and b. This lets us do arithmetic more quickly, in practice. The following proof guarantees this works, in all cases.)

Proof. Suppose $a \equiv r \bmod n$ and $b \equiv s \bmod n$. This means $\exists k, \ell \in \mathbb{Z}$ such that

$$a = kn + r$$
$$b = \ell n + s$$

Adding these equations yields

$$a + b = (kn + r) + (\ell n + s) = (k + \ell)n + (r + s)$$

Thus, $a + b \equiv r + s \bmod n$, since we can express $a + b$ as a multiple of n plus the remainder $r + s$.

Multiplying the two equations yields

$$a \cdot b = (kn + r) \cdot (\ell n + s) = k\ell n^2 + (ks + \ell r)n + r \cdot s = n \cdot \big(k\ell n + ks + \ell r\big) + r \cdot s$$

Thus, $a \cdot b \equiv r \cdot s \bmod n$, since we can express $a \cdot b$ as a multiple of n plus the remainder $r \cdot s$. □

Remark 6.5.11. Notice that we don't mention **subtraction** or **division** here, only addition and multiplication. There are two different reasons for this. The first reason is that subtraction is just "adding a negative". Thus, the lemma does actually say we can *subtract* two congruences by applying two steps: (1) multiply one of the congruences by -1 (invoking the lemma for *multiplication*),

and (2) add the results (invoking the lemma for *addition*). Notice how it uses *both* of the results of the lemma. Neat, right?

The second reason is slightly more complicated. There is really no such thing as "dividing" modulo n. The main reason is that we are restricting our discussion here to the *integers*, and division might result in *rational numbers* that are not integers. For instance, we know $4 \equiv 7 \bmod 3$, but does that tell us that $\frac{4}{2} \equiv \frac{7}{2} \bmod 3$? What does that even mean? How is an integer (namely, 2) possibly congruent to a non-integer (namely, 7/2)? For this reason, mainly, we don't discuss the operation of **division** in the context of $\mathbb{Z}$ modulo n.

There are more subtle details to this "division" issue, too, and we will have occasion to discuss them later in Section 6.5.3, when we talk about *multiplicative inverses*. For the sake of avoiding confusion now, we will not attempt to discuss those details. Suffice it to say, though, we will develop something that feels "a lot like" division modulo n, but it will only be possible in particular situations.

In the meantime, to make sure we are only talking about *integers*, we will stick to addition and multiplication *only*.

Two Examples of Usefulness

We're not sure yet whether we've convinced you that any of this modular arithmetic is even *useful* or helpful. To make sure we've established that these notions of congruence as an equivalence relation are both mathematically interesting *and* applicable, we are going to consider here two interesting and useful examples. The first is just a simply stated problem that is vastly easier to solve using modular arithmetic than "standard" arithmetic. The second is a handy trick that we're sure you've used before, but you might have never considered *how* or *why* it works. We'll prove it!

Example 6.5.12. Consider the following problem:

> **Questions:**
>
> Does there *exist* a natural number k such that 5^k is 1 more than a multiple of 7?
>
> If so, what is the *smallest* such natural number?
>
> Can you characterize *all* of the natural numbers with this property?

We might try to answer these questions by just plugging in values for k and seeing what happens. However, you'll quickly notice that computing large exponential numbers can be cumbersome, and figuring out whether a large number is exactly one more than some multiple of 7 is even harder! Go ahead, try it if you'd like. Use a calculator if you want, even. See if you can solve it!

Here's what we'd rather do, though: let's take advantage of the Modular Arithmetic Lemma (MAL) over and over. Exponentiation is just repeated multiplication, so let's just invoke the multiplication result of the lemma over and over. The idea is that we can keep multiplying by 5 and *reduce everything modulo 7 along the way*. That is, we only need to find a number that is 1 more than

a multiple of 7—i.e. congruent to 1 modulo 7—and we don't need to know immediately what *exactly* that number is, only whether or not it *has that property.* Let's show you what we mean.

We start with $5^1 \equiv 5 \mod 7$. We multiply this by 5, yielding

$$5^2 \equiv 5 \cdot 5 \equiv 25 \equiv 4 \mod 7$$

We found this by just noticing that $25 = 21 + 4$, and knowing 21 is a multiple of 7. (When the numbers are "small" like this, we can often do arithmetic **by inspection**. That is, we can just look at it for a minute and do some mental arithmetic. Of course, we could always apply the Division Algorithm if we weren't sure, just subtracting 7s from 25 until we were left with a remainder.)

We can then find

$$5^3 \equiv 5^2 \cdot 5 \equiv 4 \cdot 5 \equiv 20 \equiv 6 \mod 7$$

Again, we just found " by inspection" that $20 = 14+6$. Notice that we now know what 5^3 is congruent to modulo 7 but we *didn't have to actually compute* $5^3 = 125$ *and then reduce it.* Because we have been reducing all the numbers modulo 7 along the way, we are saving ourselves a lot of computation. Specifically, we are always reducing to something *smaller* than 7, so the largest numbers we might even have to look at, in any case, are in the 20s and 30s. How convenient! Let's keep going and see what we get:

$$5^4 \equiv 5^3 \cdot 5 \equiv 6 \cdot 5 \equiv 30 \equiv 2 \mod 7$$

$$5^5 \equiv 5^4 \cdot 5 \equiv 2 \cdot 5 \equiv 10 \equiv 3 \mod 7$$

$$5^6 \equiv 5^5 \cdot 5 \equiv 3 \cdot 5 \equiv 15 \equiv 1 \mod 7$$

That's what we were looking for! We have ascertained that 5^6 is 1 more than a multiple of 7. And this was **much easier** than calculating $5^6 = 15625$ and figuring out that $15625 = 7 \cdot 2232 + 1$, wasn't it?

This has answered the first two questions: we have found that there *exists* a power of 5 with the desired property, and since we found it iteratively (starting from $k = 1$), we know that this is the *smallest* such number. We will leave it to you to investigate the third question of characterizing *all* such numbers. Try continuing our process, multiplying by 5s and reducing. Do you notice a pattern? What is it? Make a conjecture. Try to prove it! (We'll come back to this example later on . . .)

Example 6.5.13. Consider the number 474. Is it a multiple of 3 or not? Perhaps you just added up its digits—$15 = 4 + 7 + 4$—and noticed that 15 is a multiple of 3, and then concluded that 474 must *also* be a multiple of 3. (Of course, you could also have just done the long division to find $474 = 3 \cdot 158$.) Why can you do that? Is it because your teachers told you about this in 3rd grade and you took their word for it? That's not good enough for us! ☺

Here, we will formally **prove** that a natural number x is divisible by 3 if and only if the sum of its digits is also divisible by 3. (Within the proof, we

have included parenthetical statements that work out the details with a specific *example*. We've included them to help you understand what we're writing, but we've put them in parentheses to remind you that simply showing an example is *not* a formal proof. It can help a reader understand the *actual* proof more easily, but an example alone is not enough to prove this *universally-quantified* statement.)

Proof. Let $x \in \mathbb{N}$ be arbitrary and fixed. We can represent this number using its decimal expansion by writing

$$x = \sum_{k=0}^{n-1} x_k \cdot 10^k$$

where n is the number of digits in the number x, and x_k is the digit corresponding to the 10^k-th place, so $0 \leq x_k \leq 9$. (That is, x_k is the $(k+1)$-th digit of x reading right to left.)

(For example, we can write 47205 as $47205 = 4{\cdot}10^4 + 7{\cdot}10^3 + 2{\cdot}10^2 + 0{\cdot}10^1 + 5{\cdot}10^0$. In this case, $x_0 = 5$ and $x_1 = 0$ and $x_3 = 2$ and so on.)

The Divisiblity Trick claims that

$$x \equiv 0 \mod 3 \iff \sum_{k=0}^{n-1} x_k \equiv 0 \mod 3$$

To prove this, we will consider the decimal expansion modulo 3. Notice that $10 \equiv 1 \bmod 3$, since $10 = 9 + 1$. Thus,

$$\forall k \in \mathbb{N} \cup \{0\}\,.\; 10^k \equiv 1^k \equiv 1 \mod 3$$

(This follow from the Modular Arithmetic Lemma and the fact that $1^k = 1$ for any k. Think about this!)

This allows us to replace the powers of 10 in the decimal expansion with 1s! Therefore,

$$\begin{aligned}
x \equiv 0 \mod 3 \quad &\iff \quad \sum_{k=0}^{n-1} x_k \cdot 10^k \equiv 0 \mod 3 && \text{Rewrite } x \text{ in decimal form} \\
&\iff \quad \sum_{k=0}^{n-1} x_k \cdot 1^k \equiv 0 \mod 3 && \text{Since } 10 \equiv 1 \mod 3 \\
&\iff \quad \sum_{k=0}^{n-1} x_k \equiv 0 \mod 3
\end{aligned}$$

This proves the claim.

□

(Notice that $3 \mid 47205$ because $3 \mid (4 + 7 + 2 + 0 + 5)$, which is to say $3 \mid 18$. In fact, $15735 \cdot 3 = 47205$).

Interestingly enough, we have actually proved a **stronger** result here. Because the statements above are *if and only if* statements, we actually know something further: if the sum of the digits of x is *not* a multiple of 3, then x is *not* a multiple of 3 and has the *same remainder*. For example, $3 \nmid 122$ because $3 \nmid 5$; furthermore, $5 \equiv 2 \bmod 3$ and so we know that $122 \equiv 2 \bmod 3$. (Indeed, $122 = 3 \cdot 40 + 2$.)

One can find and prove similar divisiblity tricks for 9 and 11 (although the one for 11 is a little trickier). There is even one for 7, but it's difficult to write down. These ideas will be explored in this chapter's exercises.

Remember this result and its proof. It's a good one to whip out at a party. Challenge your friends: do they actually know **why** this trick works? You do!

6.5.2 Equivalence Classes modulo n

You proved (see Lemma 6.5.9) that congruence modulo n is, indeed, an equivalence relation on the underlying set $\mathbb{Z}$. You also proved (see Theorem 6.4.10) that the equivalence classes of an equivalence relation *partition* the underlying set. Combining these two results, we know that congruence modulo n yields a partition of $\mathbb{Z}$. How convenient! How will we represent those equivalence classes, though? What would be a *natural* choice as a representative of each class?

Let's start with two simpler questions: (1) How *many* equivalence classes are there of $\mathbb{Z}$ modulo n? (2) How "*big*" are the equivalence classes?

How many equivalence classes?

To answer question (1), we just have to remember how we defined *remainders* upon division by n. The Division Algorithm (see Lemma 6.5.2) required a remainder r to satisfy $0 \leq r \leq n - 1$, when we divide some other number by n. This indicates that there are at most n possibilities for what a remainder could be: either 0 or 1 or 2 or ... or $n - 1$. (That is, $r \in [n - 1] \cup \{0\}$.) Are we sure that there *exist* numbers whose remainders are these possibilities? Sure, we can just use those numbers themselves! Certainly $n - 1$ has remainder $n - 1$ when we divide it by n (since $n - 1 < n$). Together, these observations tell us there are **exactly** n **equivalence classes** of $\mathbb{Z}$ modulo n.

We can also identify some natural choices for *representatives* of these equivalence classes via these same observations! Since $a \equiv b \bmod n$ means that a and b have the same remainder upon division by n, why don't we just declare that those two numbers belong to the equivalence class represented by *that remainder*, whatever it is. That remainder r must satisfy $0 \leq r \leq n - 1$, and we will write $a, b \in [r]_{\bmod n}$ to indicate that a and b belong to that equivalence class represented by the remainder r (with the subscript "mod n", as well, to indicate the remainder comes from division by n).

How big are the classes?

Let's think of this by using a particular value, say $n = 4$. What does it mean for an integer $z \in \mathbb{Z}$ to belong to the equivalence class corresponding to 0? That is, if we know $z \in [0]_{\text{mod } 4}$, what can we say about z?

By the definition of mod, we know this means z has a remainder of 0 when we divide by 4. Aha! This means z is a *multiple* of 4. How many multiples of 4 are there in $\mathbb{Z}$? Infinitely many! We have $4, 8, 12, 16, \ldots$, as well as $0, -4, -8, -12, \ldots$. The set $[0]_{\text{mod } 4}$ is an *infinite* set.

What about knowing $z \in [1]_{\text{mod } 4}$? What does this say about z? Having a remainder of 1 means z is expressible as $4k + 1$; that is, there *exists* such a k that allows us to write z in this way. What could that k be? Well, *any* choice of $k \in \mathbb{Z}$ creates such a number z, so we can consider letting $k = 0$ and $k = 1$ and $k = 2 \ldots$ as well as $k = -1$ and $k = -2 \ldots$ and see what happens. We find that this generates the set

$$\begin{aligned}[1]_{\text{mod } 4} &= \{\ldots, -7, -3, 1, 5, 9, \ldots\} \\ &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = 4k + 1\} = \{4k + 1 \mid k \in \mathbb{Z}\}\end{aligned}$$

Notice that we resorted to "..." notation at first to show you the pattern we noticed, and then we rewrote this set using set-builder notation (in two different ways).

This is also an infinite set. We will let you play around with other remainders (upon division by 4, as well as for general n), to let you discover that these sets are all *infinite*. (Also, we have not provided a proper, *formal* definition of what it means for a set to be *infinite* yet, but we are relying on our collective intuition for what this means. If you're looking for a good way to think about it, try this: this set is infinite because we can start to list all its elements, and identify a pattern that we are sure *will* generate all its elements, but this process will not *end* in a finite amount of time.)

The Partition of $\mathbb{Z}$ modulo n

Let's take these observations we've made about the equivalence classes and use them to make a summary about a *canonical* (i.e. standard/natural/convenient) representation of the equivalence classes of $\mathbb{Z}$ modulo n. We know there are n equivalence classes, each of them infinitely big. We know that each class corresponds to *exactly* one of the remainders you might get when dividing an integer by n. Since that remainder must satisfy $0 \leq r \leq n - 1$, we will use the set $\{0, 1, 2, \ldots, n - 1\} = [n - 1] \cup \{0\}$ as the set of canonical representatives.

The equivalence class corrseponding to remainder r will collect together all the integers which yield that remainder when divided by n. Said another way, all of the elements $z \in [r]_{\text{mod } n}$ must be exactly r more than some multiple of n. This means we can *generate* all the elements of the equivalence class by starting with r and adding/subtracting n over and over and over. (Think about it, and you'll realize this means that any two elements of the same equivalence class differ by a multiple of n.)

The equivalence classes of $\mathbb{Z}$ modulo n:

Given $n \in \mathbb{N}$, there are exactly n equivalence classes:

$$[0]_{\text{mod } n},\ [1]_{\text{mod } n},\ [2]_{\text{mod } n},\ \ldots,\ [n-1]_{\text{mod } n}$$

They are characterized by:

$$\begin{aligned}
[0]_{\text{mod } n} &= \{\ldots, -2n, -n, 0, n, 2n, \ldots\} \\
&= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = kn\} \\
[1]_{\text{mod } n} &= \{\ldots, -2n+1, -n+1, 1, n+1, 2n+1, \ldots\} \\
&= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = kn+1\} \\
[2]_{\text{mod } n} &= \{\ldots, -2n+2, -n+2, 2, n+2, 2n+2, \ldots\} \\
&= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = kn+2\} \\
&\vdots \\
[n-1]_{\text{mod } n} &= \{\ldots, -n-1, -1, n-1, 2n-1, 3n-1, \ldots\} \\
&= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = kn+(n-1)\} \\
&= \{z \in \mathbb{Z} \mid \exists \ell \in \mathbb{Z}.\ z = \ell n-1\}
\end{aligned}$$

This summarizes all of our observations in full *generality*. Here are a few examples with *specific* values of n.

- Consider $n = 2$. The equivalence classes are

$$\begin{aligned}
[0]_{\text{mod } 2} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = 2k\} = \{\text{even integers}\} \\
&= \{\ldots, -6, -4, -2, 0, 2, 4, 6 \ldots\} \\
[1]_{\text{mod } 2} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = 2k+1\} = \{\text{odd integers}\} \\
&= \{\ldots, -5, -3, -1, 1, 3, 5, \ldots\}
\end{aligned}$$

- Consider $n = 3$. The equivalence classes are

$$\begin{aligned}
[0]_{\text{mod } 3} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = 3k\} = \{\text{multiples of 3}\} \\
&= \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\} \\
[1]_{\text{mod } 3} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = 3k+1\} = \{\text{multiples of 3, plus 1}\} \\
&= \{\ldots, -8, -5, -2, 1, 4, 7, 10, \ldots\} \\
[2]_{\text{mod } 3} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}.\ z = 3k+2\} = \{\text{multiples of 3, plus 2}\} \\
&= \{\ldots, -7, -4, -1, 2, 5, 8, 11 \ldots\}
\end{aligned}$$

- Consider $n = 4$. The equivalence classes are

$$\begin{aligned}
[0]_{\text{mod } 4} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = 4k\} = \{\text{multiples of } 4\} \\
&= \{\ldots, -12, -8, -4, 0, 4, 8, 12, \ldots\} \\
[1]_{\text{mod } 4} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = 4k + 1\} = \{\text{multiples of } 4, \text{ plus } 1\} \\
&= \{\ldots, -11, -7, -3, 1, 5, 9, 13, \ldots\} \\
[2]_{\text{mod } 4} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = 4k + 2\} = \{\text{multiples of } 4, \text{ plus } 2\} \\
&= \{\ldots, -10, -6, -2, 2, 4, 6, 10, 14, \ldots\} \\
[3]_{\text{mod } 4} &= \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = 4k + 3\} = \{\text{multiples of } 4, \text{ plus } 3\} \\
&= \{\ldots, -9, -5, -1, 3, 7, 11, 15, \ldots\}
\end{aligned}$$

Using the Equivalence Classes

Why is this helpful? Why have we bothered to take you through this development of the set of integers modulo a particular equivalence relation?

The fact that $\mathbb{Z}$ is **partitioned** by these equivalence classes is extremely important. Because of this, whenever we do arithmetic in the context of $\mathbb{Z}$ modulo n, we only need to consider the equivalence classes, the **remainders**. We can reduce everything to just the numbers $\{0, 1, 2, \ldots, n-1\}$ along the way because they represent *all* the integers. We don't need to do a bunch of arithmetic with large numbers and *then* find remainders; we can just work with the remainders *alone*. Let's see a couple of examples to show you how this partition is, indeed, useful.

Example 6.5.14. Consider the following claim:

$$\forall n \in \mathbb{N}\text{.}\ 6 \mid n^3 + 5n$$

We asked you to prove this by induction on n previously! (See Problem 5.7.15 in Section 5.7.) Here, we will take advantage of equivalence classes to prove this instead!

Consider $\mathbb{Z}$ modulo 6. Since $\mathbb{N} \subseteq \mathbb{Z}$, we know that every $n \in \mathbb{N}$ must fall into **exactly one** of the equivalence classes—$[0]_{\text{mod } 6}$, $[1]_{\text{mod } 6}$, $[2]_{\text{mod } 6}$, $[3]_{\text{mod } 6}$, $[4]_{\text{mod } 6}$, $[5]_{\text{mod } 6}$—based on its remainder upon division by 6.

We can examine each case separately. Supposing n belongs to a particular equivalence class allows us to compute what equivalence class $n^3 + 5n$ must belong to. In each case, to multiply (and exponentiate, which is repeated multiplication) and add, we are applying the Modular Arithmetic Lemma 6.5.10.

(1) $n \equiv 0 \bmod 6 \implies n^3 + 5n \equiv 0^3 + 5 \cdot 0 \equiv 0 \bmod 6$

(1) $n \equiv 1 \bmod 6 \implies n^3 + 5n \equiv 1^3 + 5 \cdot 1 \equiv 6 \equiv 0 \bmod 6$

(1) $n \equiv 2 \bmod 6 \implies n^3 + 5n \equiv 2^3 + 5 \cdot 2 \equiv 18 \equiv 0 \bmod 6$

(1) $n \equiv 3 \bmod 6 \implies n^3 + 5n \equiv 3^3 + 5 \cdot 3 \equiv 42 \equiv 0 \bmod 6$

(1) $n \equiv 4 \bmod 6 \implies n^3 + 5n \equiv 4^3 + 5 \cdot 4 \equiv 84 \equiv 0 \bmod 6$

(1) $n \equiv 5 \bmod 6 \implies n^3 + 5n \equiv 5^3 + 5 \cdot 5 \equiv 150 \equiv 0 \bmod 6$

In each case, we find that $n^3 + 5n$ is a multiple of 6 (since it has a remainder of 0 when divided by 6). This tells us that $6 \mid n^3 + 5n$, no matter what n is. This proves the claim holds for all $n \in \mathbb{N}$, without using any inductive argument!

Example 6.5.15. **Quadratic Residues**:

In this example, we will investigate perfect squares. Specifically, we will look at what remainders perfect squares yield when divided by various numbers. This example will be interesting because you will notice some different patterns appear, depending on what number we are dividing by, and you might be tempted to go off and explore these patterns on your own. (If so, wonderful!) But this example will also be helpful in that some of our investigations will lead us to other results, proven in this text and in the exercises. Particularly, these investigations of perfect squares can be helpful when exploring **Pythagorean Triples**; these are triplets of integers $(a, b, c) \in \mathbb{N}^3$ that satisfy $a^2 + b^2 = c^2$. Knowing information about perfect squares can help us prove some interesting facts about these triples!

For each of the following cases, we will fix a particular $n \in \mathbb{N}$ and then investigate what x^2 reduces to modulo n, for every $x \in \mathbb{Z}$. Knowing the partition of $\mathbb{Z}$ modulo n, we can simply look at all of the n possible remainders modulo n and square them, and then reduce. These possible remainders are called the **quadratic residues** (*quadratic* because we use perfect squares, and *residues* because we find remainders). After each case, we will summarize with a list of these possible quadratic residues.

n = 2: We know that a perfect square is even if and only if the base is even, and that a perfect square is odd if and only if the base is odd. We investigated these claims back in Chapter 4 when we discussed biconditional statements and quantifiers and proof techniques. There's no need to go back and reprove those claims formally now; we can see these results easily using modular arithmetic!

Let $x \in \mathbb{Z}$ be arbitrary and fixed.

- First, suppose $x \equiv 0 \bmod 2$ (i.e. x is even). Then applying the MAL tells us $x^2 \equiv 0^2 \equiv 0 \bmod 2$ (i.e. x^2 is even).
- Second, suppose $x \equiv 1 \bmod 2$ (i.e. x is odd). Then applying the MAL tells us $x^2 \equiv 1^2 \equiv 1 \bmod 2$ (i.e. x^2 is odd).

That's it! The partition of $\mathbb{Z}$ modulo 2 tells us these are the only cases that need consideration.

Quadratic residues modulo 2: $\{0, 1\}$

n = 3: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 3 \implies x^2 \equiv 0^2 \equiv 0 \bmod 3$
- $x \equiv 1 \bmod 2 \implies x^2 \equiv 1^2 \equiv 1 \bmod 3$
- $x \equiv 2 \bmod 3 \implies x^2 \equiv 2^2 \equiv 4 \equiv 1 \bmod 3$

Quadratic residues modulo 3: $\{0, 1\}$

n = 4: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 4 \implies x^2 \equiv 0^2 \equiv 0 \bmod 4$
- $x \equiv 1 \bmod 4 \implies x^2 \equiv 1^2 \equiv 1 \bmod 4$
- $x \equiv 2 \bmod 4 \implies x^2 \equiv 2^2 \equiv 4 \equiv 0 \bmod 4$
- $x \equiv 3 \bmod 4 \implies x^2 \equiv 3^2 \equiv 9 \equiv 1 \bmod 4$

Quadratic residues modulo 4: $\{0, 1\}$

n = 5: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 5 \implies x^2 \equiv 0^2 \equiv 0 \bmod 3$
- $x \equiv 1 \bmod 5 \implies x^2 \equiv 1^2 \equiv 1 \bmod 3$
- $x \equiv 2 \bmod 5 \implies x^2 \equiv 2^2 \equiv 4 \bmod 5$
- $x \equiv 3 \bmod 5 \implies x^2 \equiv 3^2 \equiv 9 \equiv 4 \bmod 5$
- $x \equiv 4 \bmod 5 \implies x^2 \equiv 4^2 \equiv 16 \equiv 1 \bmod 5$

Quadratic residues modulo 5: $\{0, 1, 4\}$

n = 6: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 6 \implies x^2 \equiv 0^2 \equiv 0 \bmod 6$
- $x \equiv 1 \bmod 6 \implies x^2 \equiv 1^2 \equiv 1 \bmod 6$
- $x \equiv 2 \bmod 6 \implies x^2 \equiv 2^2 \equiv 4 \bmod 6$
- $x \equiv 3 \bmod 6 \implies x^2 \equiv 3^2 \equiv 9 \equiv 3 \bmod 6$
- $x \equiv 4 \bmod 6 \implies x^2 \equiv 4^2 \equiv 16 \equiv 4 \bmod 6$
- $x \equiv 5 \bmod 6 \implies x^2 \equiv 5^2 \equiv 25 \equiv 1 \bmod 6$

Quadratic residues modulo 6: $\{0, 1, 3, 4\}$

n = 7: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 7 \implies x^2 \equiv 0^2 \equiv 0 \bmod 7$
- $x \equiv 1 \bmod 7 \implies x^2 \equiv 1^2 \equiv 1 \bmod 7$
- $x \equiv 2 \bmod 7 \implies x^2 \equiv 2^2 \equiv 4 \bmod 7$
- $x \equiv 3 \bmod 7 \implies x^2 \equiv 3^2 \equiv 9 \equiv 2 \bmod 7$
- $x \equiv 4 \bmod 7 \implies x^2 \equiv 4^2 \equiv 16 \equiv 2 \bmod 7$

- $x \equiv 5 \bmod 7 \implies x^2 \equiv 5^2 \equiv 25 \equiv 4 \bmod 7$
- $x \equiv 6 \bmod 7 \implies x^2 \equiv 6^2 \equiv 36 \equiv 1 \bmod 7$

Quadratic residues modulo 7: $\{0, 1, 2, 4\}$

n = 8: Let $x \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $x \equiv 0 \bmod 8 \implies x^2 \equiv 0^2 \equiv 0 \bmod 8$
- $x \equiv 1 \bmod 8 \implies x^2 \equiv 1^2 \equiv 1 \bmod 8$
- $x \equiv 2 \bmod 8 \implies x^2 \equiv 2^2 \equiv 4 \bmod 8$
- $x \equiv 3 \bmod 8 \implies x^2 \equiv 3^2 \equiv 9 \equiv 1 \bmod 8$
- $x \equiv 4 \bmod 8 \implies x^2 \equiv 4^2 \equiv 16 \equiv 0 \bmod 8$
- $x \equiv 5 \bmod 8 \implies x^2 \equiv 5^2 \equiv 25 \equiv 1 \bmod 8$
- $x \equiv 6 \bmod 8 \implies x^2 \equiv 6^2 \equiv 36 \equiv 4 \bmod 8$
- $x \equiv 7 \bmod 8 \implies x^2 \equiv 7^2 \equiv 49 \equiv 1 \bmod 8$

Quadratic residues modulo 8: $\{0, 1, 4\}$

We will let you go on and investigate other quadratic residues. You could even try to write a computer program that will generate these lists for you. Do you notice any patterns? Given $n \in \mathbb{N}$, how many quadratic residues are there modulo n? What are they? Can you *guarantee* certain numbers that *do* and *do not* appear in any given list? Try and explore!

Example 6.5.16. Let's generalize the idea of the previous example and look at some *cubic residues*, as applied to a particular situation.

Suppose $x, y, z \in \mathbb{Z}$ satisfy $x^3 + y^3 = z^3$.

Prove that at least one of the values $\{x, y, z\}$ is a multiple of 7.

Restating our goal, we want to prove that

$$x \equiv 0 \mod 7 \ \vee \ y \equiv 0 \mod 7 \ \vee \ z \equiv 0 \mod 7$$

To do this, let's examine what the cubic residues are modulo 7.

Let $z \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $z \equiv 0 \bmod 7 \implies z^3 \equiv 0^3 \equiv 0 \bmod 7$
- $z \equiv 1 \bmod 7 \implies z^3 \equiv 1^3 \equiv 1 \bmod 7$
- $z \equiv 2 \bmod 7 \implies z^3 \equiv 2^3 \equiv 8 \equiv 1 \bmod 7$
- $z \equiv 3 \bmod 7 \implies z^3 \equiv 3^3 \equiv 9 \cdot 3 \equiv 2 \cdot 3 \equiv 6 \bmod 7$
- $z \equiv 4 \bmod 7 \implies z^3 \equiv 4^3 \equiv 16 \cdot 4 \equiv 2 \cdot 4 \equiv 8 \equiv 1 \bmod 7$

- $z \equiv 5 \bmod 7 \implies z^3 \equiv 5^3 \equiv 25 \cdot 5 \equiv 4 \cdot 5 \equiv 20 \equiv 6 \bmod 7$
- $z \equiv 6 \bmod 7 \implies z^3 \equiv 6^3 \equiv (-1)^3 \equiv -1 \equiv 6 \bmod 7$

(Notice that we conveniently chose to write 6 as -1 modulo 7 to make the calculations easier.)

We see that the only possibilities are $\{0, 1, 6\}$.

Now, suppose we have a solution to the equation, i.e. we have $x, y, z \in \mathbb{Z}$ that satisfy $x^3 + y^3 = z^3$. Each term—x^3, y^3, z^3—is congruent to either 0 or 1 or 6 modulo 7. Let's look at some cases.

- Suppose $x^3 \equiv 0 \bmod 7$. Then y^3 can satisfy any of the other possibilities—i.e. y^3 can be congruent to 0 or 1 or 6 modulo 7—and we just require z^3 to fall into the same equivalence class.

  No matter what, in this case we have $x^3 \equiv 0 \bmod 7$.

- Suppose $y^3 \equiv 0 \bmod 7$. The same argument we just used applies to x^3 and z^3, but no matter what, we have $y^3 \equiv 0 \bmod 7$.

- Suppose $x^3 \equiv 1 \bmod 7$.

  AFSOC $y^3 \equiv 1 \bmod 7$. Then $x^3 + y^3 \equiv 1 + 1 \equiv 2 \bmod 7$, but this is not possible because 2 is not a cubic residue modulo 7.

  However, we see that $y^3 \equiv 0 \bmod 7$ is a possibility, because we could have $x^3 + y^3 \equiv 1 + 0 \equiv 1 \bmod 7$.

  Also, we see that $y^3 \equiv 6 \bmod 7$ is a possibility, because we could have $x^3 + y^3 \equiv 1 + 6 \equiv 7 \equiv 0 \bmod 7$.

  No matter what, in this case we have *at least* one of the cubes—either y^3 or z^3—congruent to 0 modulo 7.

- Suppose $y^3 \equiv 1 \bmod 7$. The same argument we just used applies to x^3 and z^3, so we find that, no matter what, at least one of the cubes is congruent to 0 modulo 7.

- Suppose that $x \overset{\equiv}{} 6 \bmod 7$.

  AFSOC $y^3 \equiv 6 \bmod 7$. Then $x^3 + y^3 \equiv 6 + 6 \equiv 12 \equiv 5 \bmod 7$, but this is not possible because 5 is not a cubic residue modulo 7.

  However, we see that $y^3 \equiv 0 \bmod 7$ is a possibility, because we could have $x^3 + y^3 \equiv 6 + 0 \equiv 6 \bmod 7$.

  Also, we see that $y^3 \equiv 1 \bmod 7$ is a possibility, because we could have $x^3 + y^3 \equiv 6 + 1 \equiv 7 \equiv 0 \bmod 7$.

  No matter what, in this case we have *at least* one of the cubes—either y^3 or z^3—congruent to 0 modulo 7.

- Again, supposing $y^3 \equiv 6 \bmod 7$, the same argument applies to x^3 and z^3.

We have now seen that, no matter what situation applies, there is *at least one* cube that is congruent to 0 modulo 7. The cube which has this property depends on the situation (and in some cases, more than one cube has this property), but there is always at least one.

This is fruitful for us, because we can look back at our list of cubic residues and notice something: the *only* base whose cube is congruent to 0 modulo 7 is 0 itself! Said another way,

$$\forall z \in \mathbb{Z}\text{.}\ z^3 \equiv 0 \mod 7 \implies z \equiv 0 \mod 7$$

This means that, in every situation outlined above, we have at least one of the cubes being congruent to 0 modulo 7, which means further that we have at least one of the *base* variables congruent to 0 modulo 7. By listing the possibilities and then analyzing a few cases, we have now proved a property about *all possible solutions* to this equation without having to find any solutions!

Now, with all of that work already done, we have some unfortunate news: the *only* solution to the original equation is the *trivial* one, where $x = y = z = 0$. That's it! You can try to find other solutions, but your efforts will be in vain. This fact is a particular instance of the conclusion of **Fermat's Last Theorem**, which says that non-trivial integral solutions (i.e. where $x, y, z \in \mathbb{Z}$) exist to the equation $x^k + y^k = z^k$ (where $k \in \mathbb{N}$) if and only if $k = 1$ or $k = 2$; that is, the only solution when $k \in \mathbb{N} - \{1, 2\}$ is $x = y = z = 0$.

This fact was stated by Fermat himself when he was alive, but he never published a proof. He claimed—in the margins of one of his notebooks—to have a short proof that would not fit inside those margins, but we have come to realize that this was probably not true. Although Fermat worked in the 1600s, this theorem was only proven in the 1990s! Furthermore, the proof involved a lot of powerful mathematics that was developed over the time between Fermat's statement and the eventual proof.

If we know this theorem, then we can prove the statement in this example quite easily! If the only solution is $x = y = z = 0$, then obviously at least one of the values is a multiple of 7; they all are! This is no fun, though, and does not give us any practice working with modular arithmetic and equivalence classes.

Example 6.5.17. Here is another problem that deals with cubic residues:

> Suppose $x, y, z \in \mathbb{Z}$ satisfy $x^3 + y^3 + z^3 = 3$.
>
> Prove that $x^3 \equiv y^3 \equiv z^3 \mod 9$.

This problem concerns a particular *Diophantine Equation*. This is a general term for these types of equations that involve polynomials with multiple variables and coefficients that are integers. A *solution* to such a Diophantine Equation is a selection of values for the variables that are *integers* that satisfy the equation. Here, we are saying that *any* solution to this equation must make all of the terms—x^3 and y^3 and z^3—congruent modulo 9.

To start off, try finding a couple of solutions to this equation, just to see some examples. We'll give you a few easy ones to get you started: we might

have (x, y, z) equal to $(1, 1, 1)$ or $(4, 4, -5)$. Do you see that these solutions have the specified property? Can you find any other solutions? (This is a difficult problem, so don't try *too* hard.)

Interestingly enough, we can prove this claim without even trying to identify what all of the solutions "look like" or trying to find them. All we have to do is find what the cubic residues are modulo 9:

Let $z \in \mathbb{Z}$ be arbitrary and fixed. Applying MAL tells us:

- $z \equiv 0 \bmod 9 \implies z^3 \equiv 0^3 \equiv 0 \bmod 9$
- $z \equiv 1 \bmod 9 \implies z^3 \equiv 1^3 \equiv 1 \bmod 9$
- $z \equiv 2 \bmod 9 \implies z^3 \equiv 2^3 \equiv 8 \bmod 9$
- $z \equiv 3 \bmod 9 \implies z^3 \equiv 3^3 \equiv 9 \cdot 3 \equiv 0 \bmod 9$
- $z \equiv 4 \bmod 9 \implies z^3 \equiv 4^3 \equiv 16 \cdot 4 \equiv (-2) \cdot 4 \equiv -8 \equiv 1 \bmod 9$
- $z \equiv 5 \bmod 9 \implies z^3 \equiv 5^3 \equiv 25 \cdot 5 \equiv (-2) \cdot 5 \equiv -10 \equiv 8 \bmod 9$
- $z \equiv 6 \bmod 9 \implies z^3 \equiv 6^3 \equiv 36 \cdot 6 \equiv 0 \cdot 6 \equiv 0 \bmod 9$
- $z \equiv 7 \bmod 9 \implies z^3 \equiv 7^3 \equiv 49 \cdot 7 \equiv 4 \cdot (-2) \equiv -8 \equiv 1 \bmod 9$
- $z \equiv 8 \bmod 9 \implies z^3 \equiv 8^3 \equiv (-1)^3 \cdot -1 \equiv 8 \bmod 9$

Notice that, in some cases, we used *negative* numbers to make the calculations easier. This is totally fine, and can be helpful for you! For instance, rather than computing $4^3 = 64$ and then trying to reduce modulo 9, we can replace 16 with -2 to keep the numbers small. We can always add or subtract a multiple of 9 from any term, so we might as well try to do this along the way, instead of finding a large number and then reducing it modulo 9. (Of course, this point may seem totally moot because 64 isn't that large of a number; however, this is far more relevant when you have to work with bigger numbers. Furthermore, doing this reduction to single digits, whenever possible, can count down on the prevalence of mental arithmetic errors!) Notice that we only saw three possibilities on the far right-hand sides; the cubic residues modulo 9 are $\{0, 1, 8\}$. That's it!

Certainly, to make $x^3 + y^3 + z^3 = 3$—an *equality*—we definitely need $x^3 + y^3 + z^3 \equiv 3 \bmod 9$, since $3 \equiv 3 \bmod 9$. But looking at the sums of the possible cubic residues—0 and 1 and 8—we see that $1 + 1 + 1$ is the *only* sum that yields 3. Try the others: $0 + 1 + 8 \equiv 9 \equiv 0 \bmod 9$ and $8 + 8 + 8 \equiv 24 \equiv 6 \bmod 3$ and so on. This means that we *require* $x^3 \equiv y^3 \equiv z^3 \equiv 1 \bmod 9$ for (x, y, z) to be a solution.

In solving this, we have proven a slightly strongly result. Not only do we now know that x^3, y^3, z^3 must be congruent modulo 9, they must be congruent to 1 modulo 9. This is a little bit more information than was required of us.

Now, it turns out that something even *stronger* is true about this problem. It happens to be the case that $x \equiv y \equiv z \bmod 9$. That is, not only are the *cubes* congruent modulo 9, the *bases* are, as well. (Notice this doesn't say the bases

are congruent to 1 modulo 9; in fact, our other example of $(4, 4, -5)$ shows this doesn't have to be the case.) Unfortunately, proving this fact delves into a lot of higher mathematics, and falls far outside the scope of this book. This should give you some appreciation, though, for the idea that such "simple" problems (easily stated, small numbers, integers) require incredibly complex and deep mathematics to be solved. Rather than see this as a discouragement, though, think of it as an inspiration: with only a bit of mathematical knowledge, we could scrape the surface of this problem which hints further at very profound and intricate underpinnings.

(If you are curious, here is a paper that solves the full result, proving that $x \equiv y \equiv z \bmod 9$, necessarily:

http://www.ams.org/journals/mcom/1985-44-169/
S0025-5718-1985-0771049-4/S0025-5718-1985-0771049-4.pdf

You will have to look up some definitions to read even the first two paragraphs. This will also require you to learn the corresponding mathematics, which will probably take, oh . . . a few months or years, perhaps, depending on your interest. Keep it in mind, and bo back to it later on in your mathematical career!)

6.5.3 Multiplicative Inverses

We mentioned before—when we proved the MAL, Lemma 6.5.10—that we weren't going to talk about "division" in the context of $\mathbb{Z}$ modulo n. In this section, we will revisit that idea and explain why (and how) there are some "nice" situtations in which "division" makes sense. However, we want to stress that we are actually appealing to a more general idea of **multiplicative inverses**, and that we should **not** actually be thinking of this in terms of "division". We will explain this first with a couple of motivating examples, and then we will state and prove a result about exactly what these "nice" situations are.

The General Concept

Given a particular mathematical objects, its **multiplicative inverse** is another object such that when we "multiply" the two objects together, we get "1". We are using scare quotes here because the notions of "multiply" and "1" depend greatly on the context!

Example 6.5.18. Let's consider a familiar example first. Suppose our context is the set of real numbers $\mathbb{R}$ with the usual multiplication. Let's take the number 2. What is its multiplicative inverse? That is, is there another real number x such that $2 \cdot x = 1$, and if so, what is it? Certainly, $x = \frac{1}{2}$ works! Notice that $2 \cdot \frac{1}{2} = 1$. For this reason, we write

$$2^{-1} = \frac{1}{2} \quad \text{in the context of } \mathbb{R}$$

When we "divide both sides of an equation by 2", we are actually *multiplying* both sides by the *multiplicative inverse* of 2.

Example 6.5.19. Let's consider a perhaps less familiar example now. Consider a wall clock, with notches for the 12 hours equally spaced around its rim. We will consider rotating the clock around, so let's declare that the standard placement—with 12 at the top—is our "1". That is, this is the *usual* representation with no extra rotation, so let's call this our *identity*, our *unit element*. In essence, our "1" is the clock after the "0° rotation".

Now, let's say that "multiplying" two rotations together is simply doing one rotation after the other. For example, let's say we rotate the clock around (clockwise, of course) by 45°, and then we rotate the clock even further (clockwise) by another 90°. In our context, we have just *multiplied* the objects "45° rotation" and "90° rotation". This has produced another object, the "135° rotation".

The point of establishing these conventions—what our context is, what the objects are, what "multiply" means, and what "1" means—is that we can identify the *multiplicative inverse* of any rotation. If you think about it for a minute, you'll see that if we take the object "θ (in degrees) rotation" and *multiply* it by the "$360 - \theta$ (in degrees) rotation", then we have rotated the clock completely around by 360° and arrived with the standard placement, our "1" in this context. This means

$$\big(\theta \text{ (in degrees) rotation}\big)^{-1} = 360 - \theta \text{ (in degrees) rotation}$$

in our current context.

These two examples are meant to show you that the idea of an *inverse* is a general idea, and is not tied to any standard context of *dividing* numbers. In fact, we will see another example of this idea later on, when we talk about the *inverse of a function*. (In that context, "multiplication" is the composition of functions, and "1" is the identity function. You'll see what we mean later on in the next chapter, but we wanted to point this out now, in case you are already familiar with these concepts.)

Relatively Prime Integers

You might be familiar with the following definition. We will use it in the forthcoming result that declares when multiplicative inverses exist (in the context of $\mathbb{Z}$ modulo n), so we want to reiterate it for you now and show you some examples.

Definition 6.5.20. *Given $x, y \in \mathbb{Z}$, we say x and y are* **relatively prime** *if and only if they have no common factors (divisors), other than 1.*

(**Note:** The phrase "relatively prime" means x and y are relatively prime *to each other*. It does *not* say that x is "kinda prime-like" or anything like that.)

Example 6.5.21. For example, 12 and 35 are relatively prime, because $12 = 2^2 \cdot 3$ and $35 = 5 \cdot 7$, so we can see that they don't have any common factors.

It helps, in general, to write out these *prime factorizations* because we are really wondering whether two numbers have any *prime* factors in common (which

would imply they have a factor in common.)

For a non-example, 12 and 33 are not relatively prime, because $3 \mid 12$ and $3 \mid 33$.

Example 6.5.22. This example will be helpful to have in hand later on, after the result stated below.

Claim: If p is a prime and a is an integer that is not a *multiple* of p, then p and a are relatively prime.

(That is, if p is prime and $p \nmid a$, then p and a are relatively prime.)

Let's see why this is true!

Proof. Let p be a prime and let $a \in \mathbb{Z}$. Suppose $p \nmid a$.

Since $p \nmid a$, then *none* of the prime factors of a are p. Since p is prime itself, then none of those prime factors of a divide p, either. This means that a and p don't share *any* prime factors, so they are relatively prime. □

This is convenient! In particular, we now know that whenever p is a prime, **all** of the numbers $1, 2, 3, \ldots, p-1$ are relatively prime to p.

Definition and Examples

Let's talk about what *multiplicative inverse* means in the context of $\mathbb{Z}$ modulo n. Here, "multiply" means the usual multiplication, but everything is reduced modulo n. Also, "1" really means the *equivalence class* corresponding to 1. In this sense, we will say that for any $x \in \mathbb{Z}$, its multiplicative inverse—written as x^{-1}—is equal to y if and only if $xy \equiv 1 \mod n$. That is,

$$\forall x \in \mathbb{Z}.\ \forall y \in \mathbb{Z}.\ y \equiv x^{-1} \mod n \iff xy \equiv 1 \mod n$$

Notice that all of these claims are made in the context of $\mathbb{Z}$ modulo n, so we don't write "$y = x^{-1}$". The number x represents an entire equivalence class, as does x^{-1}.

Let's practice *finding* these multiplicative inverses, or determining when they don't exist. The key observation to make here is the following:

$$\text{If } x \cdot y \equiv 1 \mod n, \text{ then } x \cdot (y + kn) \equiv 1 \mod n \text{ for every } k \in \mathbb{Z}.$$

To see why, we can just distribute the x in the expression on the right:

$$x \cdot (y + kn) \equiv xy + xkn \equiv xy + n(xk) \equiv xy + 0 \equiv xy \equiv 1 \mod n$$

That is, adding a multiple of n to y will just yield a multiple of n in the expansion, and we can "throw it away" when we reduce everything modulo n.

The consequence of this fact is this: **If** x has a multiplicative inverse modulo n, **then** (a) there are *infinitely* many such inverses and they all belong to the same equivalence class modulo n, but (b) we can find exactly *one* such inverse in the set $\{1, 2, 3, \ldots, n-1\}$.

These facts are helpful and interesting. In particular, this tells us that we don't have to make some crazy or complicated existence argument to try and find multiplicative inverses: we can simply check the cases one by one until we find one. If we don't, then none exist. Put another way, we don't have to "intuit" the answer or randomly guess-and-check; we have a more methodical guess-and-check algorithm.

Let's see this in practice with the following examples.

Example 6.5.23. Throughout this example, we will provide an $n \in \mathbb{N}$ and an $x \in \mathbb{Z}$, and we will seek a y that satisfies $y \equiv x^{-1} \bmod n$. If no such inverse exists, we will show why.

- **$n = 3$ and $x = 2$**:

  We know we just need to check $y = 1$ and $y = 2$. Notice that $2 \cdot 2 \equiv 4 \equiv 1 \mod 3$, so

  $$2^{-1} \equiv 2 \mod 3$$

- **$n = 4$ and $x = 3$**:

  We know we just need to check $y = 1$ and $y = 2$ and $y = 3$. Notice that $3 \cdot 3 \equiv 4 \equiv 1 \mod 3$, so

  $$3^{-1} \equiv 3 \mod 4$$

- **$n = 4$ and $x = 2$**:

  We know we just need to check $y = 1$ and $y = 2$ and $y = 3$. However, notice that x is even, so any multiple of x is also even, yet any number $y \equiv 1 \mod 4$ must be odd. Thus, 2 has *no* multiplicative inverse modulo 4.

- **$n = 10$ and $x = 3$**:

  We can just check the cases here:

  $$\begin{aligned} 3 \cdot 1 &\equiv 3 \mod 10 \\ 3 \cdot 2 &\equiv 6 \mod 10 \\ 3 \cdot 3 &\equiv 9 \mod 10 \\ 3 \cdot 4 \equiv 12 &\equiv 2 \mod 10 \\ 3 \cdot 5 \equiv 15 &\equiv 5 \mod 10 \\ 3 \cdot 6 \equiv 18 &\equiv 8 \mod 10 \\ 3 \cdot 7 \equiv 21 &\equiv 1 \mod 10 \end{aligned}$$

  Aha! This means

  $$3^{-1} \equiv 7 \mod 10$$

  Notice that this also shows

  $$7^{-1} \equiv 3 \mod 10$$

because multiplication is commutative (i.e. the order does not matter).

This observation leads us to the following fact:

$(a^{-1})^{-1} \equiv a \mod n$, assuming a^{-1} exists in the first place.

- **$n = 15$ and $x = 7$**:

  If we start checking all the multiples of 7, we find that when we get to 13, we've succeeded:

  $$7 \cdot 13 \equiv 91 \equiv 6 \cdot 15 + 1 \equiv 1 \mod 15$$

  so

  $$7^{-1} \equiv 13 \mod 15$$

  We will also leave it to you to verify that, for example, 6 has *no* multiplicative inverse modulo 15.

When Do Multiplicative Inverses Exist?

Now that we've played around with a few examples, we should settle down and characterize *all* of the situations wherein multiplicative inverses exist. The following lemma does exactly this.

Lemma 6.5.24 (Multiplicative Inverses when Relatively Prime, or the MIRP Lemma)**.** *Suppose $n \in \mathbb{N}$ and $a \in \mathbb{Z}$, and that a and n are* **relatively prime***. Consider the congruence $a \cdot x \equiv 1 \mod n$. Then there exists a solution $x \in \mathbb{Z}$ to this congruence.*

In fact, there are infinitely-many solutions to this congruence, and they are all congruent modulo n. This implies there is exactly one solution in the set $[n-1] = \{1, 2, \ldots, n-1\}$.

We use a^{-1} to denote the equivalence class corresponding to the solutions of this congruence, and we call this the **multiplicative inverse** *of a modulo n.*

Furthermore, this is an if and only if *statement; that is, if a and n are* not *relatively prime, then there is* **no** *solution $x \in \mathbb{Z}$ to the congruence $a \cdot x \equiv 1 \mod n$.*

This Lemma completely characterizes when multiplicative inverses exist and when they do not. We can use it take a congruence like

$$15x \equiv 1 \mod 33$$

and declare immediately that there is **no** solution $x \in \mathbb{Z}$ because $3 \mid 15$ and $3 \mid 33$ so they are not relatively prime. Likewise, we can take a congruence like

$$40x \equiv 1 \mod 51$$

and know that there **must** be a solution $x \in \mathbb{Z}$ because $40 = 2^3 \cdot 5$ and $51 = 3 \cdot 17$ are relatively prime. (Notice that the Lemma only goes so far in helping us

find the solution; it just guarantees we can find it amongst the elements of $\{1, 2, \ldots, n-1\}$.)

To **prove** this lemma, we will split it into two parts, since it is a biconditional. We will prove one of the directions for you; namely, we will show that whenever a and n are relatively prime, a^{-1} exists modulo n. We will guide you through a proof of the other direction (if a and n share a common factor, then a^{-1} does not exist modulo n) in Problem 6.7.21. (Try to prove it now!)

We will need the following helpful lemma in our proof.

Lemma 6.5.25 (Euclid's Lemma)**.** *Let $a, b, c \in \mathbb{Z}$ be given. Suppose $a \mid bc$, and suppose a and b are relatively prime. Then $a \mid c$.*

We are going to hold off on proving this particular lemma until *after* we see the proof of the MIRP Lemma. We think that working through all the details of this proof will temporarily distract us from the main goal of this section. Also, we think that this result, Euclid's Lemma, is believable enough on its own that we can just assume it's validity for the moment and use it in the proof of the MIRP Lemma. Just look at some examples:

- We know $3 \mid 30$, and $30 = 5 \cdot 6$. Since 3 and 5 are relatively prime, we deduce that $3 \mid 6$, and it certainly does.
- Suppose $3 \mid 5x$, for some integer x. What can we say about x? Again, 3 and 5 are relatively prime, so for the product $5x$ to be a multiple of 3, it has to be the case that x "contains" a factor of 3. That is, $3 \mid x$ is a necessity.

Now, we realize this is not good enough! We are not saying that we should just *accept* this statement without proof; we just want to wait a few minutes before diving into it. In the meantime, you might want to try and prove it on your own! See what you can come up with.

Instead, let's stride ahead and prove the MIRP Lemma now (assuming the result of Euclid's Lemma, which will be used exactly once, somewhere in the middle).

Proof. Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Suppose that a and n are **relatively prime**.

WWTS $\exists x \in \mathbb{Z}\,.\; ax \equiv 1 \mod n$.

Consider the set of the first n multiples of a; that is, define the set N to be

$$\begin{aligned} N &= \{0, a, 2a, 3a, \ldots, (n-1)a\} \\ &= \{z \in \mathbb{Z} \mid \exists k \in [n-1] \cup \{0\}\,.\; z = ka\} \end{aligned}$$

Notice that there are n elements in the set N.

Claim: The elements of N all yield *distinct* remainders modulo n; that is,

$$\forall i, j \in [n-1] \cup \{0\}\,.\; i \neq j \implies ai \not\equiv aj \mod n$$

Let's prove this claim. To do so, we AFSOC the claim is False.

This means $\exists i, j \in [n-1] \cup \{0\}\text{.}\ ai \equiv aj \bmod n$. Let such i, j be given.

Subtracting and factoring tells us $ai - aj \equiv a(i-j) \equiv 0 \bmod n$.

This means $n \mid a(i-j)$. We know n and a are relatively prime. By Lemma 6.5.25 above, we can deduce that $n \mid i-j$.

Now, we claim that this implies $i = j$. Remember that $i, j \in [n-1] \cup \{0\}$, so we know that $0 \leq i, j \leq n-1$ and, thus, also $-(n-1) \leq -j, ij \leq 0$.

Adding these inequalities for i and $-j$, we find that

$$-(n-1) + 0 = n - 1 \leq i + (-j) = i - j \leq n - 1 = (n-1) + 0$$

That is, $-(n-1) \leq i - j \leq n-1$. We also know already that $n \mid i-j$, i.e. $i-j$ is a *multiple* of n. Notice, though, that the *only* multiple of n that lies between $-(n-1)$ and $+(n-1)$ is 0.

Thus, $i - j = 0$, and so $i = j$. This proves the current claim.

We now know that the elements of N yield *distinct* remainders modulo n. We also know already that those possible remainders are $\{0, 1, 2, \ldots, n-1\} = [n-1] \cup \{0\}$. Notice that there are n distinct elements of N, and there are n distinct remainders (i.e. equivalence classes) modulo n. This means that *every* remainder modulo n is represented *exactly once* in the set N.

This tells us there is *exactly* one element of N (i.e. exactly one multiple of a) that corresponds to a remainder of 1 modulo n. This element of N is of the form ax, for some $x \in [n-1] \cup \{0\}$. Let such an x be given. This is the solution to the congruence stated in the claim of the lemma. □

Phew! It took a bit of work, but now we're here. Since you have proven the claim in Problem 6.7.21 (you have, right? ☺), we now know *exactly* when multiplicative inverses exist in the context of $\mathbb{Z}$ modulo n. We also know a reasonable way of finding them: we just need to check the first $n-1$ multiples of a, looking for one that yields 1 modulo n.

Now that we have accomplished this, let's step back and prove Euclid's Lemma. This needs to be done, since the important MIRP Lemma's proof depends on this result. Notice that there is a tricky *induction* argument in this proof. Specifically, we have *two variables*—a and b—and we want to prove a certain statement holds for every such a and b. To do this,

Proof. Let $a, b, c \in \mathbb{Z}$. Suppose $a \mid bc$, and that a and b are relatively prime.

WWTS that $a \mid c$, necessarily. We will accomplish this by first proving:

Claim: If $a, b \in \mathbb{N}$ and a and b are relatively prime, then $\exists x, y \in \mathbb{Z}\text{.}\ ax + by = 1$.

From this claim, the result will follow quite easily. We have outlined the proof of this claim in a box, for ease of reading. After the box, you will find how we

use this result to prove the original statement of the lemma.

(Before working through this proof, try working with examples to "convince" yourself this claim is True. Take two relatively prime numbers—like 5 and 11, or 15 and 22, or 10 and 23—and try to construct these *linear combinations* that yield 1. Then, take some numbers that share a common divisor—like 5 and 10, or 6 and 15, or 21 and 27—and try to realize why you *cannot* find such a combination.)

Proof of claim: We will prove this by induction on the sum $a + b$. Before starting this, let us observe a few facts:

- If $a = 1$ or $a = -1$, then b must be 0 or 1 for them to be relatively prime.

  In either case, we can use $x = a$ and $y = 0$ to write

  $$ax + by = a^2 + 0 = 1$$

  The same argument applies for the case when $b = 1$ or $b = -1$ (i.e. a must be 0 or 1).

- If $b = 0$, and $|a| \geq 2$ (i.e. $a \neq 1$ and $a \neq -1$), then a and b share the common factor of a, so they are not relatively prime.

  The same argument applies for the case when $a = 0$.

Together, these observations imply that we don't need to consider either value being 0. That is, we will only consider values that satisfy $|a| \geq 1$ and $|b| \geq 1$.

- Since a and b are relatively prime, then $-a$ and $-b$ are also relatively prime (as are $-a$ and b, and as are a and $-b$). This is because negating an integer does not affect what its divisors are, only its sign.

- If we already know that $\exists x, y \in \mathbb{Z}\text{.}\ ax + by = 1$, then certainly

  $$(-a)(-x) + (-b)(-y) = ax + by = 1$$

  Since $-x, -y \in \mathbb{Z}$, as well, this shows that $-a$ and $-b$ have such a representation.

Together, these observations imply that we only need to consider values of a and b that are *positive*. (That is, if either or both are negative, we can just negate them.)

Combining this with the previous deduction, we deduce that we only need to consider $a, b \in \mathbb{N}$. Proving the result for these values will *imply* the full result, when combined with the observations we have just made.

Now, we may proceed with our proof by (strong) induction on the sum $a + b$. Since $a, b \in \mathbb{N}$, we have $a + b \geq 2$. We considered the base case $a + b = 2$ above, but we will restate it again here for completeness.

Given $a, b \in \mathbb{N}$, define $P(a, b)$ to be the statement

$$\text{“ } a \text{ and } b \text{ are relatively prime } \implies \exists x, y \in \mathbb{Z}\text{.}\ ax + by = 1 \text{ ”}$$

BC: Consider $P(2)$, i.e. assume $a, b \in \mathbb{N}$, a and b are relatively prime, and $a + b = 2$. This means $a = b = 1$, and so we can choose $x = 1$ and $y = 0$, yielding

$$ax + by = 1 + 0 = 1$$

Thus, $P(2)$ holds.

IH: Let $k \in \mathbb{N}$ be arbitrary and fixed. Suppose $P(2) \wedge P(3) \wedge \cdots \wedge P(k)$ holds. (That is, suppose that whenever two relatively prime numbers add up to 2 or 3 or ... or k, we know we can find a *linear combination* of them that makes 1.)

IS: WWTS that $P(k + 1)$ holds. That is, let $a, b \in \mathbb{N}$ be given with $a + b = k + 1$, and suppose a and b are relatively prime; WWTS that $\exists x, y \in \mathbb{Z}\text{.}\ ax + by = 1$.

First, we may assume that $a \geq b$, by symmetry. (That is, we are *given* these values a and b. Whatever they are, we can just “rename” them, because one of them must be at least as large as the other; whichever one is larger, we will label it as a.) In fact, since a and a are not relatively prime (when $a \geq 2$), we may even assume that $a > b$.

Now, we want to appeal to the fact that b and $a - b$ are *also* relatively prime. To see why this is true, we need to show that b and $a - b$ have no common divisors except 1.

Let d be a common divisor of b and $a - b$, i.e. $d \mid b$ and $d \mid a - b$. We know that this implies $d \mid b + (a - b)$, i.e. $d \mid a$. We already knew $d \mid b$, as well, so d is actually a common divisor of a and b, so it must be 1. Thus, b and $a - b$ are relatively prime.

(What we just proved there is:

$$\big(d \mid b \wedge d \mid a - b\big) \implies \big(d \mid a \wedge d \mid b\big)$$

This claim is also an $\iff$ statement, and we encourage you to think about that for a minute and see why the $\impliedby$ direction also holds.)

What we now have is $b, a - b \in \mathbb{N}$ (since $b < a$) that are relatively prime. Notice, as well, that $b + (a - b) = a < a + b = k + 1$, since $b \in \mathbb{N}$ (so $b \geq 1$). This means that $a + b \leq k$, and so the Inductive Hypothesis $P(a + b)$ applies!

(Notice that $P(a+b)$ is not necessarily $P(k)$, so we *needed* to use Strong Induction here!)

That statement $P(a+b)$—i.e. $P(b+(a-b))$, as we will use it—tells us there is a linear combination of b and $a-b$ that yields 1; that is,

$$\exists u, v \in \mathbb{Z}\text{.}\ ub + v(a-b) = 1$$

We now want to manipulate this into a linear combination of a and b that yields 1. To do this, we will just rewrite the above equation and relabel the coefficients:

$$ub + v(a-b) = 1 \iff \underbrace{v}_{x} a + b\underbrace{(u-v)}_{y} = 1$$

That is, we can now *define* $x = v$ and $y = u - v$, so that $x, y \in \mathbb{Z}$ and $ax + by = 1$.

We have now shown that $P(a+b)$ (i.e. $P(k+1)$) holds. By Strong Induction, we deduce that $P(n)$ holds for every $n \in \mathbb{N}$ with $n \geq 2$.

What the proof of this claim has accomplished, to remind you, is that we now know any relatively prime numbers can be put into a linear combination of 1.

Let's return to the original statement of the lemma. We are given $a, b, c \in \mathbb{N}$, and we supposed a and b are relatively prime and $a \mid bc$.

That first assumption tells us $\exists x, y \in \mathbb{Z}\text{.}\ ax + by = 1$. Let such x, y be given.

That second assumption tells us $\exists k \in \mathbb{Z}\text{.}\ bc = ak$. Let such a k be given.

We will multiply through that first assumption's equation by c, and then apply the second assumption:

$$ax + by = 1 \implies acx + (bc)y = 1 \implies acx + (ak)y = c \implies c = a\underbrace{\big(cx + ky\big)}_{\ell}$$

That is, knowing $ax + by = 1$ lets us deduce that $c = a\ell$, where $\ell \in \mathbb{Z}$ is defined in terms of other integers.

By definition, this means $a \mid c$. This proves the original statement. □

Wow! There was a lot going on in that proof. Make sure you read through it a few times, following along each line and making notes. Do you see why every claim follows from what we already know? Do you see how the induction worked? We had two variables to work with, but we were inducting on *one* variable, defined as the sum of the other two. We realize this is a tricky proof, which is why we put it here, after the more important result of this section, the MIRP Lemma.

Let's take this result we now have—knowing *precisely* when a multiplicative inverse exists—and use it to solve some problems!

Using Multiplicative Inverses

How is this useful? You might consider this answer to be a little cheeky, but it is certainly valid: multiplicative inverses are useful in solving congruences using modular arithmetic. Now, that might seem like we developed some mathematical tools to solve the very problems they arose from, but that's not quite the case. In fact, as you will see from the forthcoming examples, in trying to *solve* these problems, you would likely have to innovate the very techniques we will apply. That is to say, you could try to solve these problems without having studied multiplicative inverses before, but in doing so and considering more general problems, you would wind up rediscovering the results we have worked through with you!

Alright, enough preamble. Let's pose and solve a couple of problems. These are all of the form: "Here is a proposed congruence; identify all the integral solutions, or prove that there are no solutions."

Example 6.5.26. Find all integers $x, y \in \mathbb{Z}$ that satisfy

$$3x - 7y = 11$$

We claim that there are infinitely many pairs $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ that satisfy this equation. Furthermore, we can state the form of all solutions; we will do so by defining the *set* of all such solutions.

By rewriting the given equation, we see that we want to find *all* $x, y \in \mathbb{Z}$ such that

$$3x = 7y + 11$$

Put even another way, we want to find *all* $x \in \mathbb{Z}$ such that

$$3x \equiv 11 \mod 7$$

Assuming we can find all of these integers $x \in \mathbb{Z}$, we can easily find the corresonding $y \in \mathbb{Z}$ by rearranging the equation above: $y = \frac{3x-11}{7}$.

Notice that $3^{-1} \equiv 5 \bmod 7$, since $3 \cdot 5 \equiv 15 \equiv 2 \cdot 7 + 1 \equiv 1 \bmod 7$. Thus, by the MAL, we can multiply both sides of a congruence by 3^{-1}, yielding

$$\begin{aligned} \forall x \in \mathbb{Z}\text{.}\ 3x \equiv 11 \mod 7 &\iff 3^{-1} \cdot 3 \cdot x \equiv 3^{-1} \cdot 11 \mod 7 \\ &\iff 1 \cdot x \equiv 5 \cdot 4 \mod 7 \\ &\iff x \equiv 20 \equiv 6 \mod 7 \end{aligned}$$

Since we know that 3^{-1} characterizes *all* of the solutions to this congruence (i.e. it represents the equivalence class of the multiplicative inverse of 3 modulo 7), then we can deduce that

$$\forall x \in \mathbb{Z}\text{.}\ 3x \equiv 11 \mod 7 \iff x \equiv 6 \mod 7 \iff \exists k \in \mathbb{Z}\text{.}\ x = 7k + 6$$

This characterizes all of the possible values of $x \in \mathbb{Z}$ for a solution to the given equation.

Now, we use this to identify the corresponding values of $y \in \mathbb{Z}$ for a solution. Suppose we have been given $k \in \mathbb{Z}$ with $x = 7k + 6$. Then we substitute and find that

$$y = \frac{3x - 11}{7} = \frac{3\big(7k + 6\big) - 11}{7} = \frac{21k + 7}{7} = 3k + 1$$

Now, we have a form that represents *all possible solutions* to the given equation. We know that *any* $k \in \mathbb{Z}$ yields a corresponding x, which yields a corresponding y. Furthermore, since our derivation uses $\iff$ statements, we know that this characterizes *all* the solutions.

We can state the set of solutions S to the given equation by setting

$$S = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ (x, y) = (7k + 6, 3k + 1)\}$$

Interesting Fact:

In this example, we solved a **Linear Diophantine Equation** and established all of its solutions. By *linear*, we meany that the variables x and y are both only raised to the first power. There are no squares or cubes or what have you.

Using the technique we applied in this example, you can go off and solve *any* such Linear Diophantine Equation, or determine easily whether it has any solutions. In fact, we will *prove* a result about when such an equation has *no* solutions (see Bézout's Identity, Theorem 6.5.31) and the method used here applies whenever there *are* solutions.

In the next example we will look at a **Quadratic Diophantine Equation**, where the variables are squared (we will have an x^2 and y^2 term). We will talk about the possibility of solving those types of equations after that example.

Example 6.5.27. Let's now see an example that uses a similar procedure as the previous example (using multiplicative inverses to simplify) but also uses quadratic residues.

Claim: There are **no** integral solutions $x, y \in \mathbb{Z}$ to

$$3x^2 - 5y^2 = 1$$

Let $x, y \in \mathbb{Z}$ be given. WWTS that $3x^2 - 5y^2 = 1$ is *impossible.*

We start by rewriting the given equation as

$$3x^2 = 5y^2 + 1$$

This means, in particular, that

$$3x^2 \equiv 1 \mod 5$$

since $5y^2 \equiv 0 \mod 5$. Notice that $3^{-1} \equiv 2 \mod 5$, since $3 \cdot 2 = 6 = 5 + 1$. We can thus multiply both sides by 3^{-1} and simplify:

$$3x^2 \equiv 1 \mod 5 \iff 3^{-1} \cdot 3x^2 \equiv 3^{-1} \cdot 1 \mod 5 \iff x^2 \equiv 2 \mod 5$$

However, look back at Example 6.5.15 where we examined *quadratic residues*. We saw that the set of quadratic residues modulo 5 is $\{0, 1, 4\}$. That is, it is *not possible* to have an integer x satisfy $x^2 \equiv 2 \mod 5$. This means no integral solutions exist to the given equation.

Interesting Fact:

We stated above that we know exactly when Linear Diophantine Equations are solvable, and how to solve them. Unfortunately, we are not so lucky regarding these **Quadratic Diophantine Equations**. It's quite difficult to look at one and determine whether it is solvable or not. Even then, knowing it *is* solvable, it is quite difficult to actually solve it, too!

In fact, we are *extremely* unlucky with these Quadratic Diophantine Equations. It is known that there is **no possible computer algorithm that can input any Diophantine Equation with variables raised to the 1st and 2nd powers and output whether or not that equation has any solutions**. This fact doesn't even touch upon the idea of **how** to solve such an equation, just whether or not it **has** a solution. Wow! This fact is a form of Hilbert's Tenth Problem.

Rest assured, the Diophantine Equations we will give you here in the examples and the exercises will be ones you can analyze with the techniques we've provided. This fact we mentioned here is a broader statement about the class of all such equations, in generality.

A Little Bit of Group Theory

In this small section, we want to just point out that there are some powerful and profound principles of mathematics underlying the current topic. Alas, we don't have the time and space required to devote ourselves to developing these topics fully. In lieu of doing so, we will state a few ideas and facts here and illustrate them with examples.

The main idea we want to convey is that something special happens when we consider $\mathbb{Z}$ modulo p, where p is a **prime**. In that case, *every* number smaller than p is *relatively prime* to p, since p has no divisors except 1. This means that all of the numbers in $\{1, 2, \ldots, p-1\}$ *must* have multiplicative inverses modulo p. How convenient! This means that every equivalence class (except $[0]_{\mod p}$)

has a corresponding class of multiplicative inverses.

For example, consider $p = 5$. Notice that

$$\begin{aligned} 1^{-1} &\equiv 1 \mod 5 \\ 2^{-1} &\equiv 3 \mod 5 \\ 3^{-1} &\equiv 2 \mod 5 \\ 4^{-1} &\equiv 4 \mod 5 \end{aligned}$$

As another example, consider $p = 7$. Notice that

$$\begin{aligned} 1^{-1} &\equiv 1 \mod 7 \\ 2^{-1} &\equiv 4 \mod 7 \\ 3^{-1} &\equiv 5 \mod 7 \\ 4^{-1} &\equiv 2 \mod 7 \\ 5^{-1} &\equiv 3 \mod 7 \\ 6^{-1} &\equiv 6 \mod 7 \end{aligned}$$

Notice that all of the elements have a multiplicative inverse.

(Also, notice that these inverses are just a *permutation* of the numbers 1 through $p-1$. This is not a coincidence! Try to prove that this happens! Try to prove that there are two elements that are their own inverses—$1^{-1} \equiv 1 \mod p$ and $(p-1)^{-1} \equiv p-1 \mod p$—but each other element *cannot* be its own inverse.)

This is certainly *not* the case when we consider $\mathbb{Z}$ modulo n, where n is **composite**. In that case, we know there is some factorization of n; let's say we can write $n = ab$, for some $a, b \in \mathbb{N} - \{1\}$. Then $1 < a < n$ but a is *not* relatively prime to n (they share the common factor a), so a has *no* multiplicative inverse modulo n. In fact, all of the divisors of n (and their multiples) will have no multiplicative inverse modulo n.

For example, consider $n = 6$. Then,

$$\begin{aligned} 1^{-1} &\equiv 1 \mod 6 \\ 2^{-1} &\text{ Does Not Exist} \mod 6 \\ 3^{-1} &\text{ Does Not Exist} \mod 6 \\ 4^{-1} &\text{ Does Not Exist} \mod 6 \\ 5^{-1} &\equiv 5 \mod 6 \end{aligned}$$

Because of this distinction, the mathematical "structure" of $\mathbb{Z}$ modulo p stands out. It has some "nice" properties, and "behaves well" in some sense. These are vague terms we are using, of course, but the main idea is this: having inverses for *all* of its elements makes $\mathbb{Z}$ modulo p special. In fact, $\mathbb{Z}$ modulo p forms a kind of mathematical structure known as a **group**.

In general, heuristic terms, a group is a set of objects that can be "multiplied" together so that the multiplication is (a) commutative and (b) associative, and that all elements have multiplicative inverses. We already know that standard multiplication of integers (even in $\mathbb{Z}$ modulo n, for any n) is commutative and associative, and working $\mathbb{Z}$ modulo p (for prime p) tells us that every element has an inverse.

If you are interested in exploring these ideas further, we have included some exercises at the end of this chapter that address some of these properties. You might also look up an introductory textbook on **Abstract Algebra** or **Modern Algebra** or **Group Theory** or something like that. There is a lot of powerful and profound mathematical ideas out there, and **groups** are relevant and applicable in many areas!

6.5.4 Some Helpful Theorems

In this section, we will explore some theorems in number theory that appeal to modular arithmetic and are useful and interesting in their own rights. We will state and prove the theorems (sometimes, with your help via some exercises!) and then demonstrate their usefulness with examples.

Chinese Remainder Theorem

To motivate this theorem, we will first describe its usefulness by a story:

> General Sun Tzu has many soldiers in his regiment, and after a battle, he wants to count how many he has remaining. It would take a little too long to count them individually, so he wants to be more efficient. Conveniently, the soldiers have been trained well and can form themselves into equal-sized groups quite easily.
>
> General Sun Tzu orders the soldiers into two long rows of equal length, and finds there is one soldier left over.
>
> He then orders them to make three rings of equal size, but finds there is again one soldier left over.
>
> Finally, he orders them to make five flanks of equal size, but finds there are two soldiers left over.
>
> At this point, he thinks he has enough information. After the recent battle, he can declare that there are somewhere between 250 and 300 soldiers, in total, in this regiment. Using this piece of information, he knows *exactly* how many soliders there are.
>
> Can you determine this number? How many soldiers are there?

We will let you play around with this and see if you can figure it out. Read on for our solution, a theorem statement, and a description of a technique to solve these kinds of problems.

Read the story again. Letting x be the number of soldiers in General Sun Tzu's regiment, then the story tells us that x must satisfy the following three congruences and the following inequality:

$$\begin{aligned} x &\equiv 1 \mod 2 \\ x &\equiv 1 \mod 3 \\ x &\equiv 2 \mod 5 \\ 250 &\leq x \leq 300 \end{aligned}$$

(Do you see where these came from, based on the story?)

There are two things to consider now: (1) *Must* there exist an x that satisfies all three congruences? (2) Are there *several* such x values? Are we guaranteed that one such x also satisfies the inequality?

The **Chinese Remainder Theorem**, as stated below, will guarantee (1) the existence of infinitely-many solutions to the congruences, and (2) the existence of (at least) one solution that satisfies the given inequality. Before we state and prove the theorem, though, let's try to solve this initial problem. We will break this into a few observations and steps:

- The first congruence requires a solution x to be **odd**. This eliminates all even numbers as potential solutions. A list of potential solutions:

  $$1, \cancel{2}, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, 9, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, 15, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, 21, \cancel{22}, 23, \ldots$$

- The second congruence requires a solution to be 1 more than a multiple of 3. This eliminates any number congruent to 0 or 2 modulo 3. A list of potential solutions:

  $$1, \cancel{2}, \cancel{3}, \cancel{4}, \cancel{5}, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{10}, \cancel{11}, \cancel{12}, 13, \cancel{14}, \cancel{15}, \cancel{16}, \cancel{17}, \cancel{18}, 19, \cancel{20}, \cancel{21}, \cancel{22}, \cancel{23}, \ldots$$

- The third congruence requires a solution to be 2 more than a multiple of 5. This eliminates any number congruent to 0 or 1 or 3 or 4 modulo 5. A list of potential solutions:

  $$\cancel{1}, \cancel{2}, \cancel{3}, \cancel{4}, \cancel{5}, \cancel{6}, \textcircled{7}, \cancel{8}, \cancel{9}, \cancel{10}, \cancel{11}, \cancel{12}, \cancel{13}, \cancel{14}, \cancel{15}, \cancel{16}, \cancel{17}, \cancel{18}, \cancel{19}, \cancel{20}, \cancel{21}, \cancel{22}, \cancel{23}, \ldots$$

It looks like 7 is the only solution in sight, but how do we know there aren't others? We only looked at the first 23 potential solutions ... Can we be *guaranteed* there are no others? We will leave it to you to investigate this question right now. Try some larger numbers. Can you find any other solutions? Can you guess a pattern? Is 7 the only solution?

Now, let's be a little more clever about solving these congruences. Specifically, let's pretend we have a solution x (that satisfies all three congruences)

and see if we can deduce any more information about it. By the end of this derivation, we will have established a fact about all *possible* solutions to these congruences.

By the definition of modulo, we know there exist $k, \ell, m \in \mathbb{Z}$ such that

$$\begin{aligned} x &= 2k + 1 \\ x &= 3\ell + 1 \\ x &= 5m + 2 \end{aligned}$$

Let such k, ℓ, m be given.

Consider the first two equations. Let's try to combine them into one equation about x. Specifically, let's multiply the first one by 3 and the second by 2; this creates a $6k$ and 6ℓ term, respectively, so if we subtract the equations, we can factor appropriately. That is, we first find

$$\begin{aligned} 3x &= 6k + 3 \\ 2x &= 6\ell + 2 \end{aligned}$$

and then

$$(3x - 2x) = (6k + 3) - (6\ell + 2) \implies x = 6(k - \ell) + 1$$

Since $k, \ell \in \mathbb{Z}$ were given to us, let's just define $u = k - \ell$, so $u \in \mathbb{Z}$. Notice that this now tells us $x = 6u + 1$ or, in other words,

$$x \equiv 1 \mod 6$$

Now, we obtained this new congruence by combining the first two congruences, and it is *not* a coincidence that this congruence is written modulo 6, and that $6 = 2 \cdot 3$. You will see how this happens later on when we guide you through a proof of the forthcoming theorem!

Moving on, let's try to combine this new congruence with the third one given above. We'll employ a similar method: we'll multiply the one we just derived by 5 and the one given above by 6, so that when we subtract, we can factor out a 30. (This also shows you why the new congruence we will be deriving will be written modulo 30.) We get

$$\begin{aligned} 5x &= 30u + 5 \\ 6x &= 30m + 12 \end{aligned}$$

and then

$$(6x - 5x) = (30m + 12) - (30u + 5) \implies x = 30(m - u) + 7$$

Again, since m, u were given to us, let's just define $v = m - u$, so $v \in \mathbb{Z}$. This now tells us $x = 30v + 7$ or, in other words,

$$x \equiv 7 \mod 30$$

This final congruence was derived by combining each of the given congruences into each other, so it represents *all* of the information that those three congruences provided. We claim that this now tells us **all** of the solutions!

Firstly, this newly-derived congruence tells us that *any* solution must be congruent to 7 modulo 30. Said another way, any number with a remainder *other* than 7, when divided by 30, cannot possibly be a solution. Essentially, this combines all of our work in the three observations made above—where we crossed out potential solutions—into one statement.

Secondly, we can explain that, in fact, *any* number that is congruent to 7 modulo 30 *will* indeed be a solution. Let's see why. Let there be $n \in \mathbb{Z}$ and define $y = 30n + 7$ (i.e. we are choosing an arbitrary $y \in \mathbb{Z}$ that satisfies $y \equiv 7 \mod 30$). Notice that y satisfies:

- the first congruence, because $y = 30n + 7 = 2(15n + 3) + 1$, and so $y \equiv 1 \mod 2$;
- the second congruence, because $y = 30n + 7 = 3(10n + 2) + 1$, and so $y \equiv 1 \mod 3$;
- the third congruence, because $y = 30n + 7 = 5(6n + 1) + 2$, and so $y \equiv 2 \mod 5$.

That's it! We now know that (1) *any* solution x must satisfy $x \equiv 7 \mod 30$, and (2) any such x actually *is* a solution. Together, these statements form an $\iff$ statement, namely

$$x \text{ is a solution to all three congruences} \iff x \equiv 7 \mod 30$$

and therefore the set S of **all solutions** is given by

$$S = \{x \in \mathbb{Z} \mid x \equiv 7 \mod 30\} = \{30n + 7 \mid n \in \mathbb{Z}\}$$

Returning to the original problem statement, we just now need to factor in the given inequality. Is there a number x that satisfies $x \equiv 7 \mod 30$ and $250 \leq x \leq 300$? Why, yes, there is! We can find it by just starting with 7 and adding on multiples of 30, or guessing close to 300 and adjusting, or something reasonable like that. However you do it, you'll find that $\mathbf{x = 277}$ is the solution we sought all along. That is how many soldiers General Sun Tzu has in his regiment.

Now, for the sake of comparison, consider the following system of congruences that might have come from a similar problem statement:

$$x \equiv 3 \mod 4$$
$$x \equiv 2 \mod 6$$

Are there any solutions to this system of congruences? Does the method we used above apply here? If you play around with it—trying the "cross out bad candidates" method, or the "combine congruences" method—you'll find that

nothing works. Looking back at the system, you'll notice that this makes sense. The first congruence requires x to be 3 more than a multiple of 4; since a multiple of 4 is even, we are asking x to be *odd*. However, the second congruence requires x to be 2 more than a multiple of 6; since a multiple of 6 is also even, we are asking x to be *even*. How can a solution be both odd and even at the same time?! This is clearly not possible.

The **Chinese Remainder Theorem** tells us when there are *guaranteed* to be solutions to a system of congruences. It applies to the first problem we solved above, and it in fact tells us the ultimate result we found: that there are infinitely-many solutions and that they are all congruent modulo 30. However, it does *not* tell us there are *no* solutions to the second problem we just solved. This theorem is a *guarantee* that applies to certain situations. When we face those situations, we can make a valid statement about solutions. When we are facing a *different* situation, though, the theorem makes *no claim* about the existence of solutions. Let's see the statement now, then discuss it a little further, and then get your help to prove it (in two different ways!).

Theorem 6.5.28. *Suppose we are given a system of r-many different congruences. That is, suppose $r \in \mathbb{N}$ and we have r natural numbers, $n_1, n_2, \ldots, n_r \in \mathbb{N}$, and we also have r integers, $a_1, a_2, \ldots, a_r \in \mathbb{Z}$, and the system of congruences is given by*

$$\begin{aligned} x &\equiv a_1 \mod n_1 \\ x &\equiv a_2 \mod n_2 \\ &\vdots \\ x &\equiv a_r \mod n_r \end{aligned}$$

(Put another way, the system asks for $x \in \mathbb{Z}$ such that $\forall i \in [r] \bullet x \equiv a_i \mod n_i$.)

If *the moduli n_i are* pair-wise relatively prime*—that is, no two of the numbers n_i share any common factors, besides 1—***then** *the system of congruences has a solution.*

Furthermore, in this case, there are in fact infinitely-many *solutions, and they are all congruent modulo N, where N is defined as the product of the moduli:*

$$N = \prod_{i \in [r]} n_i$$

Notice that the main conclusion is the "**If**...**then**..." statement. Remember what we said about conditional statements like this? This theorem is offering *no statement* about what happens when two of the moduli are *not* relatively prime. Anything could happen in that case! The example we saw above had *non*-relatively prime moduli: one congruence was given modulo 4 and the other modulo 6, and 4 and 6 share the common factor of 2. However, the theorem does not say there are **no** solutions; we had to figure that out for ourselves. What if

we had changed the numbers slightly and posed the following congruences:

$$x \equiv 3 \mod 4$$
$$x \equiv 5 \mod 6$$

There *are* solutions to this system. Can you find them?

One proof of the Chinese Remainder Theorem follows the method we used to solve the problem above. Given any number of congruences in a system, we can iteratively combine one into the other, eventually ending up with one congruence whose modulus is the product of all the other moduli. How do you think one would prove that this method works? An iterative process ... aha, induction! Yes indeed, you can prove the Chinese Remainder Theorem by inducting on r, the number of congruences in the given system. This proof is outlined in Exercise 6.7.26. We like this proof because it also provides you with a *technique* for solving these types of problems, in practice.

Another proof is **constructive**. That is, it takes the information in the theorem statement and combines it to *define* a number X that is a solution (and proves this, of course). This proof is outlined in Exercise 6.7.27. We like this proof because it is, indeed, constructive; rather than proving an existence result by arguing for *why* a certain object exists, it actually *produces* it for you. However, the solution it constructs is *not* the same solution you would find by performing the "cross out bad candidates" or "combine congruences" methods. It's actually a somehat "unnatural" method to use, but it does indeed *work* without having to do any inductive process. For the sake of comparison, we encourage you to work through *both* proofs of this theorem. However, if we *had* to recommend only one, we would suggest the induction proof.

Bezout's Identity

This theorem hearkens back to our discussion of Linear Diophantine Equations. In Example 6.5.26, we solved a particular such equation, by prudently applying multiplicative inverses. In addition to showing you that method, we pointed out that there is a simple way of verifying whether such an equation even *has* a solution. This theorem characterizes precisely when a linear diophantine equation of two variables has a solution. It is known as **Bézout's Identity**, named after the 18th-century French mathematician Étienne Bézout.

Before stating the theorem, we need to provide one definition. You are likely familiar with it already, but it plays a starring role in this theorem, so we want to share it here and provide some illustrative examples.

Definition 6.5.29. *Let $a, b \in \mathbb{Z}$ be given. The* **greatest common divisor** *of a and b is denoted by* $\gcd(a, b)$ *and is defined to be the* largest *integer that divides both a and b. That is,*

$$\gcd(a, b) \mid a \ \wedge \ \gcd(a, b) \mid b$$

and

$$\forall d \in \mathbb{Z} \bullet \big(d \mid a \wedge d \mid b\big) \implies d \leq \gcd(a, b)$$

We will assume some familiarity with this idea, or at least some intuition for it. The forthcoming theorem, and its proof, will not depend too heavily on a thorough understanding of this concept. Furthermore, any exercises that refer to this definition or theorem will not ask you to have great computational powers, or presume any knowledge of this idea. Rather, consider this a component of your continuing practice with absorbing new *definitions* of mathematical concepts, of your ability to take these abstract notions and invoke them and apply them to prove further facts and develop examples and non-examples. This is an important skill! Before moving on to state and prove the theorem, let's just quickly look at some examples of this notion at work.

Example 6.5.30. In some of these cases, we will take two numbers and state their greatest common divisor. Usually, a reasonable way to actually *find* such a gcd is to find the **prime factorization** of both numbers, and combine them appropriately. That is to say, $\gcd(a, b)$ is the product of the prime factors that a and b have in common, so considering what those factors are tells us the gcd rather easily.

In some of these cases, we will make some claim about the gcd in general, and prove it (or perhaps ask you to prove it!). These will only rely on the definition we just provided above.

- Let $a = 15$ and $b = 6$. Since $a = 3 \cdot 5$ and $b = 2 \cdot 3$, we see that they share only the factor 3. Thus,
$$\gcd(6, 15) = 3$$
- Let $a = 30$ and $b = 40$. Since $a = 2 \cdot 3 \cdot 5$ and $b = 2^3 \cdot 5$, we see that they share one factor of 2 and one factor of 5. Thus,
$$\gcd(30, 40) = 10$$
- In general,
$$\gcd(a, b) = \gcd(b, a)$$
This is clearly True because any common divisor of a and b is *also* a common divisor of b and a.
- Let $a = 77$ and $b = 72$. Since $a = 7 \cdot 11$ and $b = 2^3 \cdot 3^2$, we see that they have no common prime factors. Thus,
$$\gcd(72, 77) = 1$$
- Let $a = 13$ and let $b \in \mathbb{N}$ such that $a \nmid b$. Since a is prime, and b is not a multiple of 13, then b cannot have 13 as a prime factor, and therefore,
$$\gcd(13, b) = 1$$
This means that a and b are **relatively prime**. This is a general fact:
$$a \text{ and } b \text{ are relatively prime} \iff \gcd(a, b) = 1$$

Furthermore,

$$\forall a, b \in \mathbb{N}\text{.}\ a \text{ prime} \implies \big(\gcd(a,b) = 1 \iff a \nmid b\big)$$

Now, we feel ready to state and prove **Bézout's Identity**!

Theorem 6.5.31 (Bézout's Identity)**.** *Let $a, b \in \mathbb{Z}$ be given. Define L to be the set of all* linear combinations *of a and b; that is, define*

$$L = \{z \in \mathbb{Z} \mid \exists x, y \in \mathbb{Z}\text{.}\ ax + by = z\} = \{ax + by \mid x, y \in \mathbb{Z}\}$$

Also, define M to be the set of all multipled of $\gcd(a, b)$*; that is, define*

$$M = \{z \in \mathbb{Z} \mid \exists k \in \mathbb{Z}\text{.}\ z = k \cdot \gcd(a,b)\} = \{k \cdot \gcd(a,b) \mid k \in \mathbb{Z}\}$$

Then,

$$L = M$$

That is, the Linear Diophantine Equation $ax + by = c$ has a solution if and only if *c is a* multiple *of* $\gcd(a, b)$*.*

How convenient! Notice that this theorem tells us *precisely* when a Linear Diophantine Equation—$ax + by = c$, given $a, b, c \in \mathbb{Z}$—has a solution. We only need to find $\gcd(a, b)$ and make sure $\gcd(a, b) \mid c$.

To prove this theorem, we need to prove a *set equality*, so we will use a *double-containment argument*, a strategy we have seen many times before. We will prove one of the containments for you here, and the other left as an exercise.

Proof. Let $a, b \in \mathbb{Z}$ be given. Define L and M as in the theorem statement.

First, we will prove $L \subseteq M$. Let $z \in L$ be arbitrary and fixed.

By the definition of L, we know $\exists x, y \in \mathbb{Z}\text{.}\ ax + by = z$. Let such x, y be given.

Since $\gcd(a, b)$ divides both a and b, we know $\exists k, \ell \in \mathbb{Z}$ such that $a = k \cdot \gcd(a, b)$ and $b = \ell \cdot \gcd(a, b)$. Let such k, ℓ be given.

We take these expressions for a and b and replace them in the equation above:

$$z = ax + by = k \cdot \gcd(a,b) \cdot x + \ell \cdot \gcd(a,b) \cdot y = \gcd(a,b) \cdot \underbrace{(kx + \ell y)}_{m}$$

Define $m = kx + \ell y$. Since $m \in \mathbb{Z}$, this shows that z is a multiple of $\gcd(a, b)$.

Therefore, $z \in M$. This shows that $L \subseteq M$.

Second, we must prove $M \subseteq L$ ...

This is left for the reader as Exercise 6.7.12 □

With this result now completed and proved, we know whether a Linear Diophantine Equation of two variables has a solution. Several exercises will ask you to determine whether such an equation has a solution. To do this, just cite

this result. If you are further asked to *find* all solutions, apply the techniques we showed you in Example 6.5.26

Challenge Question: What do you think can be said about Linear Diophantine Equations of *more* than two variables? For example, consider

$$6x + 8y + 15z = 10$$

Does this equation have any solutions? How many? As another example, consider

$$3x + 6y + 9z = 2$$

Does this equation have any solutions? Why or why not?

Try to state and prove a result about this. Can you generalize it to any number of variables?

6.5.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) Why is $\mathbb{Z}$ partitioned into several sets when we consider $\mathbb{Z}$ modulo n?

(2) What are the equivalence classes of $\mathbb{Z}$ modulo n?

(3) How can you determine whether two integers $x, y \in \mathbb{Z}$ belong to the same equivalence class of $\mathbb{Z}$ modulo n?

(4) What does the Modular Arithmetic Lemma say? Why is it helpful? How can we use it to manipulate congruences algebraically?

(5) What does the general concept of *multiplicative inverse* mean? Given $a \in \mathbb{Z}$ and $n \in \mathbb{N}$, how can you determine whether the multiplicative inverse of a exists, in the context of $\mathbb{Z}$ modulo n?

(6) What is special about the set of equivalence classes of $\mathbb{Z}$ modulo p when p is a *prime*?

(7) Is the following system of congruences *guaranteed* to have a solution by the Chinese Remainder Theorem? Why or why not?

$$x \equiv 2 \mod 6$$
$$x \equiv 5 \mod 9$$

Can you identify a solution to the system? (**Hint:** Yes, you can!)

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) State and prove a *divisibility trick* for determining whether a natural number $x \in \mathbb{N}$ is a multiple of 9.

(**Hint:** See Example 6.5.13 for a similar problem.)

(2) Let $n \in \mathbb{N}$ and let $a \in \mathbb{Z}$. Show that $(n-a)^2 \equiv a^2 \mod n$.

(3) Let $n \in \mathbb{N} - \{1\}$. Show that $(n-1)^{-1} \equiv n-1 \mod n$.

(4) Given each of the pairs of values (a, n), find the multiplicative inverse of a modulo n, or else say that it does not exist.

(a) $a = 5$ and $n = 12$
(b) $a = 7$ and $n = 11$
(c) $a = 6$ and $n = 27$
(d) $a = 11$ and $n = 18$
(e) $a = 70$ and $n = 84$
(f) $a = 8$ and $n = 17$

(5) Characterize all integral solutions $x, y \in \mathbb{Z}$ of the equation

$$4x - 7y = 18$$

(6) Identify *all* solutions to the following system of congruences:

$$x \equiv 3 \mod 5$$
$$x \equiv 4 \mod 7$$

6.6 Summary

As part of our ongoing buildup to formalizing **functions**, we have thoroughly discussed binary relations. We defined a relation as just a set of ordered pairs, and we talked about several properties that relations might have. In particular, the combination of *reflexivity*, *symmetry*, and *transitivity* yield a particularly powerful relationship known as an *equivalence relation*. We saw a helpful theorem about such relations, which says that an equivalence relation corresponds precisely with a *partition*. With the particular equivalence relations "modulo n" defined on $\mathbb{Z}$, we were able to take advantage of these partitions and state and prove several interesting results about the integers! Many of the exercises below address our work with abstract relations, while many of them also address our work in the area of number theory and the integers.

6.7 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them. Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ⋆.

Problem 6.7.1. ► Consider the set $A = \{1, 2, 3, 4\}$. For each of the following relations, defined on A or $\mathcal{P}(A)$ as specified, decide whether it is (i) reflexive, (ii) symmetric, (iii) transitive, (iv) anti-symmetric.

(a) R_a on A defined by $R_a = \{\ (1,2), (2,2), (3,1), (4,2), (3,3)\ \}$

(b) R_b on A defined by $R_b = \{\ (1,1), (2,2), (3,3), (3,4), (4,3), (4,4)\ \}$

(c) R_c on $\mathcal{P}(A)$ defined by $\forall S, T \in \mathcal{P}(A)\,.\ (S,T) \in R_c \iff S - T \subseteq \{1\}$

(d) R_d on $\mathcal{P}(A)$ defined by $\forall S, T \in \mathcal{P}(A)\,.\ (S,T) \in R_d \iff S \cap T \subseteq \{1\}$

Problem 6.7.2. Define the relation $\sim$ on $\mathbb{R}$ by setting

$$\forall a, b \in \mathbb{R}\,.\ a \sim b \iff \forall x \in \mathbb{R}\,.\ x > 0 \implies ax^2 + bx > 0$$

For each of the four properties of relations—(i) reflexive, (ii) symmetric, (iii) transitive, (iv) anti-symmetric—either prove that $\sim$ has that property, or else disprove it by finding a counterexample.

Problem 6.7.3. Define the relation $\approx$ on $\mathcal{P}(\mathbb{R})$ by setting

$$\forall X, Y \in \mathcal{P}(\mathbb{R})\,.\ X \approx Y \iff X - Y \subseteq \mathbb{N}$$

For each of the four properties of relations—(i) reflexive, (ii) symmetric, (iii) transitive, (iv) anti-symmetric—either prove that $\sim$ has that property, or else disprove it by finding a counterexample.

Problem 6.7.4. Define the relation $\#$ on $\mathbb{Z} \times \mathbb{N} - \{0\}$ by setting

$$\forall (a,b), (c,d) \in \mathbb{Z} \times \mathbb{N} - \{0\}\,.\ (a,b)\,\#\,(c,d) \iff ad = bc$$

(a) Prove that $\#$ is an equivalence relation.

(b) Identify the elements in the equivalence class $[(0,3)]$. Prove your claim.

(c) Identify the elements in the equivalence class $[(2,3)]$. Prove your claim.

(d) Identify the elements in the equivalence class $[(-2,2)]$. Prove your claim.

(e) How *many* equivalence classes are there in $(\mathbb{Z} \times \mathbb{N} - \{0\})/\#$?

Problem 6.7.5. Let $p \in \mathbb{N}$ be an odd prime (i.e.$p \neq 2$). Prove that $p^2 \equiv 1 \mod 24$.

Problem 6.7.6. Use Euclid's Lemma (see Lemma 6.5.25) to prove that prime factorizations of natural numbers are **unique**.

(Note: We proved previously, in Example 5.4.3, that prime factorizations *exist*, but we did not prove their uniqueness.)

Problem 6.7.7. Define the relation T on $\mathbb{R}$ by setting

$$\forall x, y \in \mathbb{R}\text{.}\ \ (x,y) \in T \iff \left(\frac{y}{x} \in \mathbb{R} \wedge \frac{y}{x} \geq 0 \right)$$

(a) For every $x \in \mathbb{R}$, let the set $S(x)$ be

$$S(x) = \{y \in \mathbb{R} \mid (x,y) \in T\}$$

Write down what the sets $S(-1)$, $S(0)$, and $S(1)$ are.

(b) Use the three sets from part (a) to deduce that T is **not** an equivalence relation.

(c) Verify this result by showing that T is not reflexive and not symmetric.

(d) Is T transitive or not? Prove your claim.

Problem 6.7.8. Consider the following *spoof*. Identify which claim in the *argument* is incorrect. Then, provide an explanation (including an **example**) as to why the *conclusion* of the argument is also incorrect.

> Let $n \in \mathbb{N}$ and let $a, b, x \in \mathbb{Z}$. Suppose that $ax \equiv bx \mod n$. We claim that we can "cancel" and deduce that $a \equiv b \mod n$.
>
> Since $ax \equiv bx \mod n$ then, by definition, $n \mid ax - bx$. Thus, $n \mid x(a-b)$, and so $n \mid a - b$. By definition, then, $a \equiv b \mod n$.

Problem 6.7.9. Consider the system of congruences given by:

$$\begin{aligned} x &\equiv 1 \mod 2 \\ x &\equiv 2 \mod 6 \end{aligned}$$

Does the **Chinese Remainder Theorem** guarantee the existence of a solution X? Can you find a solution?

Problem 6.7.10. In Definition 6.5.29, we defined the **greatest common divisor** of two integers to be the *largest* integer that divides both of them.

Here, we want you to prove that the following definition of gcd is **equivalent** to the one we provided. First, read the definition:

Definition: Let $a, b \in \mathbb{Z}$ be given. Define $G(a, b)$ to be a common divisor of a and b such that *all* common divisors of a and b divide it. That is,

$$G(a, b) \mid a \ \wedge \ G(a, b)$$

and

$$\forall d \in \mathbb{Z}\text{.} \ \big(d \mid a \ \wedge \ d \mid b\big) \implies d \mid G(a, b)$$

Now, prove that this definition is equivalent. That is, prove that

$$\forall a, b \in \mathbb{Z}\text{.} \ \gcd(a, b) = G(a, b)$$

Problem 6.7.11. Consider the following False (obviously!) claim:

Claim: 1 is a multiple of 3.

What is wrong with the following "spoof" of the claim:

> WWTS $1 \equiv 0 \bmod 3$. Observe that
>
> $$\begin{aligned} 1 \equiv 4 &\implies 2^1 \equiv 2^4 \equiv 2 \equiv 16 \implies 2 \equiv 1 \\ &\implies 2 - 1 \equiv 1 - 1 \implies 1 \equiv 0 \end{aligned}$$

Problem 6.7.12. Complete the proof of Bézout's Identity (Theorem 6.5.31) by proving that $M \subseteq L$. (These sets are defined in the theorem statement.)

Problem 6.7.13. In this problem, you will prove the converse of Theorem 6.4.12. Namely, you will prove the following: **Theorem:** Let $S \neq \varnothing$ be a set and let R be an equivalence relation on S. The set of equivalence classes S/R forms a **partition** of S.

Remember that we use the notation $[x]_R$ to mean the **equivalence class corresponding to** x, and it is the set of all elements of S that are related to x; that is,

$$[x]_R = \{y \in S \mid (x, y) \in R\}$$

Throughout the parts of this problem, we are assuming that S is a set and R is an equivalence relation on S, so that R is reflexive, symmetric, and transitive.

(a) Let $x \in S$. Show that $x \in [x]_R$.

(b) Let $x, y \in S$. Suppose $x \neq y$, and suppose that $(x, y) \in R$. Show that $[x]_R = [y]_R$.

(**Hint:** Use transitivity. You'll need it *twice*.)

(c) Let $x, y \in S$. Suppose $x \neq y$, and suppose that $(x, y) \notin R$. Show that $[x]_R \cap [y]_R = \varnothing$.

(**Hint:** Use a contradiction argument.)

(d) Explain why this has proven the stated **Theorem**.

Problem 6.7.14. In this problem, you will prove the Division Algorithm, stated in Lemma 6.5.2. That is, you will prove that

$$\forall a, b \in \mathbb{Z}\,.\; \exists!\, k, r \in \mathbb{Z}\,.\; ak + r = b \wedge 0 \leq r \leq a - 1$$

1. Let $M = \{\ell \in \mathbb{Z} \mid \ell a \leq b\}$. Prove that M has a **maximum** element.
2. Let $k \in M$ be that maximum element. Define $r = b - ka$. Prove that $0 \leq r \leq a - 1$.
3. Suppose $K, R \in \mathbb{Z}$ also satisfy $aK + R = b$ and $0 \leq R \leq a - 1$. Prove that $K = k$ and $R = r$, thereby showing k, r are *unique*.

Problem 6.7.15. Prove Lemma 6.5.8, that $n \mid a - b \iff a$ and b have the same remainder upon divisio by n

Problem 6.7.16. Prove Lemma 6.5.9. That is, prove that congruence modulo n is indeed an equivalence relation on $\mathbb{Z}$.

(**Hint:** Just prove it is (1) reflexive, (2) symmetric, and (3) transitive.)

Problem 6.7.17. This problem asks you to prove/disprove some statements about **Pythagorean Triples**, which are triplets $(x, y, z) \in \mathbb{N}^3$ that satisfy $x^2 + y^2 = z^2$.

In each case, determine whether the property must *necessarily* hold. If so, prove it. Otherwise, find a counterexample.

(a) Is it necessarily true that at least one of $\{x, y, z\}$ is even?

(b) Is it necessarily true that at least one of $\{x, y, z\}$ is a multiple of 3?

(c) Is it necessarily true that at least one of $\{x, y, z\}$ is a multiple of 4?

(d) Is it necessarily true that at least one of $\{x, y, z\}$ is a multiple of 5?

Problem 6.7.18. State and prove a *divisibility trick* for determining whether a natural number $x \in \mathbb{N}$ is a multiple of 11.

(**Hint:** See Example 6.5.13 for a similar problem.)

Problem 6.7.19. Notice that there are several "small" primes that are congruent to 3 modulo 4; for instance $3, 7, 11, 19, 23, 31 \equiv 3 \bmod 4$. In this problem, you will prove that there are, in fact, *infinitely many* primes of this form!

(You might notice that the steps of this proof closely mimic our proof that there are infinitely many prime numbers!)

(a) Suppose $n \in \mathbb{N}$ and $n \equiv 3 \bmod 4$. Prove that there must exist a prime p that satisfies $p \equiv 3 \bmod 4$ and $p \mid n$.

(**Hint:** $3 \equiv -1 \bmod 4$.)

(b) Now, AFSOC there are only *finitely* many primes that satisfy $p \equiv 3 \bmod 4$. Let's define the set of these particular primes to be $P = \{p_1, p_2, \ldots, p_k\}$, where p_k is the largest such prime.

Define the new number $N = p_1 \cdot p_2 \cdot p_3 \cdots p_k$.

Explain why N must be odd, and why N is strictly larger than all of the primes in the particular set P.

(c) Define M to be the next largest number, after N, that is congruent to 3 modulo 4. Explain why $M - N$ is either 2 or 4.

(d) Explain why, in either case ($M - N = 2$ or $M - N = 4$), it follows that *none* of the prime factors of N can be a prime factor of M.

(**Hint:** Recall that $a \mid b \wedge a \mid c \implies a \mid (b \pm c)$.)

(e) Use what you have proven so far to explain why M must be prime.

(f) What is the contradiction at which we have arrived? Make a conclusion.

Problem 6.7.20. Mimic the details of the previous Problem 6.7.19 to prove that there are also infinitely many primes that are congruent to 5 modulo 6.

Problem 6.7.21. In this problem, you will prove the second conclusion of the MIRP Lemma 6.5.24. Specifically, you will prove the following claim:

> Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}$ be given, and suppose a and n are *not* relatively prime. Then there are no solutions $x \in \mathbb{Z}$ to the congruence $ax \equiv 1 \bmod n$.

(a) We have supposed that a and n are not relatively prime. What does this imply?

(b) AFSOC that $\exists x \in \mathbb{Z}.\ ax \equiv 1 \bmod n$ and let such an x be given.

Use this to write an *equation* (not a *congruence*) involving a, x, n.

(c) Invoke your knowledge from part (a) and rewrite the equation.

(d) What contradiction have you found?

Problem 6.7.22. For each of the following claims, determine whether it is True or False. If it is True, prove it; if it is False, find a counterexample.

(a) $\forall x, y \in \mathbb{Z}.\ (x + y)^2 \equiv x^2 + y^2 \bmod 2$

(b) $\forall x, y \in \mathbb{Z}.\ (x + y)^3 \equiv x^3 + y^3 \bmod 3$

(c) $\forall x, y \in \mathbb{Z}.\ (x+y)^4 \equiv x^4 + y^4 \mod 4$

(d) $\forall x, y \in \mathbb{Z}.\ (x+y)^5 \equiv x^5 + y^5 \mod 5$

(e) $\forall x, y \in \mathbb{Z}.\ (x+y)^6 \equiv x^6 + y^6 \mod 6$

Challenge: Can you make a conjecture about which values of n make the following statement True?

$$\forall x, y \in \mathbb{Z}.\ (x+y)^n \equiv x^n + y^n \mod n$$

Can you **prove** it? Can you also characterize what values of n make the statement False?

Problem 6.7.23. Determine whether or not there are any integral solutions $x, y \in \mathbb{Z}$ to the equation

$$3x^2 - 5y^2 = 2$$

(**Hint:** Use multiplicative inverses and quadratic residues.)

Problem 6.7.24. Prove that there are no integral solutions $x, y \in \mathbb{Z}$ to the equation

$$3x^2 - 5y^2 = 15$$

Problem 6.7.25. For each of the following equations, identify the set of all integral solutions $x, y \in \mathbb{Z}$, or else explain why no such solutions exist.

(a) $2x + 4y = 9$

(b) $18x - 15y = 21$

(c) $6x - 15y = 17$

(d) $6x - 15y = 33$

Problem 6.7.26. In this problem, you will prove the Chinese Remainder Theorem (Theorem 6.5.28) by *induction*. Then, you will apply the iterative method developed in the proof to solve a particular system of congruences.

(a) Suppose we have two congruences to solve simultaneously

$$\begin{aligned} x &\equiv a_1 \mod n_1 \\ x &\equiv a_2 \mod n_2 \end{aligned}$$

where n_1, n_2 are relatively prime.

Use the definition of "modulo" to write two **equations** from these congruences. Combine these equations algebraically to deduce **one** congruence, that is written modulo $n_1 n_2$.

(b) Use the assumption that n_1, n_2 are relatively prime to deduce that $n_2 - n_1, n_1 n_2$ are also relatively prime.

(**Hint:** You will need Euclid's Lemma 6.5.25.)

(c) Deduce that you can write a single congruence $X \equiv$ _____ mod $n_1 n_2$.

This has proven the base case: that we can combine two congruences into one.

(d) Now, prove the inductive step:

Suppose $r \in \mathbb{N} - \{1\}$ and we have r natural numbers, $n_1, n_2, \ldots, n_r \in \mathbb{N}$, that are pair-wise relatively prime. (That is, no two of the numbers have any common factors, besides 1.) Suppose we also have r integers, $a_1, a_2, \ldots, a_r \in \mathbb{Z}$.

We will have you prove

$$\exists X \in \mathbb{Z}\,.\ \forall i \in [r]\,.\ X \equiv a_i \mod n_i$$

by induction on r, the number of congruences given.

Use what you proved already in this problem to rewrite this as a system of $r - 1$ congruences.

(e) Explain why this has proven the Chinese Remainder Theorem by induction.

(f) Now, consider the following system of congruences:

$$\begin{aligned} x &\equiv 2 \mod 3 \\ x &\equiv 2 \mod 5 \\ x &\equiv 4 \mod 7 \end{aligned}$$

Apply the iterative method that is generated by the above proof to solve the system.

Problem 6.7.27. In this problem, you will prove the Chinese Remainder Theorem (Theorem 6.5.28) by a *constructive* method. Then, you will apply this constructive method to solve a particular system of congruences.

Suppose $r \in \mathbb{N}$ and we have r natural numbers, $n_1, n_2, \ldots, n_r \in \mathbb{N}$, that are pair-wise relatively prime. (That is, no two of the numbers have any common factors, besides 1.) Suppose we also have r integers, $a_1, a_2, \ldots, a_r \in \mathbb{Z}$.

We will have you prove

$$\exists X \in \mathbb{Z}\,.\ \forall i \in [r]\,.\ X \equiv a_i \mod n_i$$

by helping you to define such an X and then proving it does indeed satisfy all of the congruences.

Throughout this problem, we use N as given by the definition in the theorem statement:

$$N = \prod_{i \in [r]} n_i$$

(a) For every $i \in [r]$, define $N_i = \frac{N}{n_i}$. Explain why n_i and N_i are relatively prime.

(b) Cite a result that guarantees (for every $i \in [r]$) the existence of an integer y_i that satisfies $y_i N_i \equiv 1 \mod n_i$.

(c) Define

$$X = \sum_{j=1}^{r} a_j N_j y_j$$

Our goal now is to prove that $X \equiv a_i \mod n_i$ for every $i \in [r]$.

Let $i \in [r]$ be arbitrary and fixed. Show that for every $j \neq i$, the corresponding term in the sum above is congruent to 0 modulo n_i; that is, show

$$\forall j \in [r] \,.\, j \neq i \implies a_j N_j y_j \equiv 0 \mod n_i$$

(d) Take i to be the same fixed value as in the last part. Now, show that when $j = i$, the corresponding term in the sum above that defines X is congruent to a_i modulo n_i; that is, show

$$a_i N_i y_i \equiv a_i \mod n_i$$

(e) Use what you have just proven to explain why X satisfies *all* of the r-many congruences.

Bonus Prove the second conclusion of the **CRT**, that all solutions are congruent modulo N.

(f) Now, consider the following system of congruences:

$$\begin{aligned} x &\equiv 2 \mod 3 \\ x &\equiv 2 \mod 5 \\ x &\equiv 4 \mod 7 \end{aligned}$$

That is, $n_1 = 3, n_2 = 5, n_3 = 7$, and $a_1 = 2, a_2 = 2, a_3 = 4$.

Following the definitions in the steps above, find N and N_i and y_i (for every $i \in [3]$) and use these to find a solution X.

(g) Use the other conclusion of the **CRT** to write down the set of *all* solutions to the given system in the previous part using set-builder notation, and use this to find the **smallest** natural number that is a solution.

Problem 6.7.28. The following puzzle was posed by the Indian mathematician **Brahmagupta** in the 7th century. (It just goes to show that people have been thinking about these kinds of problems for thousanda of years!)

Read it and use the story to state a system of congruences.

Then, solve the problem!

(**Hint:** We suggest some kind of iterative method, since the Chinese Remainder Theorem does not apply here, as the problem is stated. [Why not?] Could you perhaps be clever about the first steps of your method so that the Chinese Remainder Theorem *does* then apply?)

> A woman was returning from the market, carrying a basket of eggs. All of a sudden, a man walking by bumped into her, splilling the basket of eggs onto the ground. All of them broke!
>
> "I am so sorry!" said the man. "Please allow me to walk to the market and buy you more eggs to replace these. How many did you have?"
>
> The woman looked at the ground, only to find a muddled mess of shells, yelks, and mud. There would be no hope of simply counting them here.
>
> "I can't recall precisely," she said to the man, "but I do remember these facts:
>
> I first tried to count the eggs in pairs, but there was one left over. Then I counted them by threes, and there were two left over. Then I counted them by fours, and there were three left over. Then I counted them by fives, and there were four left over. Then I counted them by sixes, and there were five left over. Lastly, when I counted them by sevens, they divided evenly, so I stacked them in my basket that way. Alas, I don't recall how many groups of seven there were!"
>
> "No matter," replied the man, grinning knowingly. "You have already told me enough. I know how many eggs you had, and I shall return in a few minutes with an equal number, plus a sweetbread for your troubles." He smiled at the woman, turned around, and walked off to market.
>
> The woman stood there for a few minutes waiting, and in that time, she figured out how many eggs she had bought, as well.
>
> How many eggs were there?

Problem 6.7.29. **Challenge:** Investigating Equivalence Relations

(a) Suppose R and S are equivalence relations on the set A. Suppose that $A/R = A/S$ (i.e. the set of equivalence classes under each equivalence relation are the same). Prove that, in fact, $R = S$.

(b) Suppose R and S are equivalence relations on the set A. Must $R \cap S$ be an equivalence relation?

(c) Suppose R and S are equivalence relations on the set A. Must $R \cup S$ be an equivalence relation?

(d) Suppose R and S are equivalence relations on the set A. Define the *composition* of the relations to be

$$S \circ R = \{(x, z) \in A \times A \mid \exists y \in A \mathbf{.}\ (x, y) \in R \wedge (y, z) \in S\}$$

Must $S \circ R$ be an equivalence relation?

(e) Suppose R and S are equivalence relations on the set A. Recall that A/R and A/S are *partitions* of A.

We say that a partition $\mathcal{F}$ **refines** a partition $\mathcal{G}$ if and only if

$$\forall X \in \mathcal{F} \mathbf{.}\ \exists Y \in \mathcal{G} \mathbf{.}\ X \subseteq Y$$

Prove that

$$R \subseteq S \iff A/R \text{ refines } A/S$$

6.8 Lookahead

We are now prepared to formally discuss **functions**. We have developed the requisite background knowledge, temrinology, and notation to not only mathematically define functions, but also discuss their varied properties and prove some powerful theorems. Although it might seem like our foray into equivalence relations and number theory was purely out of intrigue and not usefulness, this is far from true! Some of the number-theoretic results we discussed will be useful in the next few chapters as we discuss some further properties of the integers and other sets. Also, we will be able to talk about functions on sets of equivalence classes, for example. Essentially, don't feel like anything we've done is a standalone result. As we are realizing, all of mathematics is connected somehow! For one, we are about to see that a **function** is just a special kind of **relation** ...

Chapter 7

Functions and Cardinality: Inputs, Outputs, and The Sizes of Sets

7.1 Introduction

We are continuing our two-chapter development of functions. In this chapter, we will formally *define* a function. Specifically, we will see that a function is really a particular kind of **relation** with certain properties. That's why we took the time to explore relations to begin with—besides the fact that they are interesting and useful in their own right, of course. After defining a function, we will explore what kinds of properties functions might have through many examples and proofs. The definitions and theorems and proofs we see in this exploration will make use of all of the concepts we have developed so far, especially the proof techniques from Section 4.9.

Later in this chapter, we will use the concept of a *bijective* function—essentially, a "pairing up" of elements of two sets—to talk about the "sizes" of sets and how to compare them. This topic, *cardinality*, will show us some rather remarkable and counterintuitive facts about infinite sets. It will also provide us an inroad into the next chapter, where we restrict our focus to finite sets and how to count them.

7.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of

statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to ...

- State the definition of a function, and provide many examples.
- Take informal descriptions and visual diagrams of functions and use them to construct formal arguments about examples (and non-examples) of functions and their properties.
- Define images and pre-images of sets, in the context of a function, and prove various properties of these operations.
- State several properties of functions, as well as apply relevant techniques to determine and prove whether or not a given function has these properties.
- Find the composition of two functions, recognize how this can be used to create new functions, and identify and prove what consequences composition has on the properties of the functions of involved.
- Describe the relationship between bijective functions and inverses, and use this to solve problems and prove claims.
- Use bijections to define the cardinalities of sets and prove claims about these cardinalities.
- State the difference between finite, countably infinite, and uncountably infinite sets, and provide several examples of each type.

7.1.2 Segue from previous chapter

The important idea from the previous chapter that will be helpful in this one is that of a **relation**. As we mentioned already, we will see that a **function** is just a particular kind of relation. This will come up in our formal definition of what a function is.

The other ideas explored in the previous chapter—equivalence relations, and results in number theory—will not appear so explicitly in this chapter. That is to say, the examples of functions we explore here, and their properties, do not depend on the other ideas explored in the last chapter. Rather, we will use those ideas to create interesting examples and exercises here.

7.1.3 Motivation

As we mentioned in the last chapter, it's quite likely that you have an intuitive understanding of what a function is and how to work with it. This might come from previous work in other mathematics courses, or perhaps some computer programming. As we have been stressing all along, we want to properly, formally, and *mathematically* define the concepts we work with. Functions are no exception! By accomplishing this, we will also be better able to talk about some of the qualitative properties of functions you may have seen before but been unable to express. As mentioned above, as well, particular properties of functions will allow us to talk about the *cardinalities* of sets. Rest assured, we would be unable to have a proper discussion of this topic *without* exploring functions first!

7.1.4 Goals and Warnings for the Reader

We want to make the same warnings and recommendations we did in the last chapter, as well. We are continuing to explore some abstract areas of mathematics. This chapter, in particular, will take a concept you might be familiar with *visually* and *intuitively* and place it on a more rigorous standing. Whenever possible, we will appeal to our collective intuitions, but there will be no way to avoid the kind of abstract thinking and problem-solving we have been developing, either. In particular, we won't always be able to associate a function with its **graph**, which is a standard (and helpful, mind you) way of learning about graphs early on in our mathemaical careers. Furthermore, some of the results presented in our discussion of *cardinality* will laugh wholeheartedly in the face of your intuitions. Seriously! We will see some strange, counter-intuitive facts, and it will help to keep an open mind about them.

7.2 Definition and Examples

How do you usually think of a function? What's your intuitive sense for what it is? How would you actually *define* it in terms of mathematical objects? Have you even tried to do so before? Do it! Think about the concepts and tools we have seen already. Is there a way you can get across the way you usually think of a *function* using just those concepts? Seriously, try it! Read the next couple of paragraphs first, as we build up to the definition, and then try to come up with one on your own.

The way we usually think of a function is that it is some kind of *rule* or *map* that tells us how to assign output values to any given input value. For example, let's take the function on $\mathbb{R}$ that says $f(x) = x^2$. This function "takes in" a real number and "spits out" the real number that is the square of the input. In a way, the function f is a "machine" that turns a number into its square; saying that the function is "on $\mathbb{R}$" means that we are only allowed to put real numbers into the machine. How do we know what's allowed to come *out* of the machine, though? Already, we have found a flaw in our usual interpretation of a function. Ideally, we would like to convey all essential information about a function when

we define it: what the inputs can be, what the outputs can be (not what all of them *are*, necessarily, but what types of objects they *could* be), and what the "rule" is. If you're thinking of a function as a *map*, then it's like a description of how to navigate from one set of numbers (in this case, $\mathbb{R}$) to another set of numbers (also $\mathbb{R}$, in this case) by following a certain "road" between the sets (in this case, taking the square of the first number). In this interpretation, we still want to convey all of the information we mentioned a moment ago, but we're just pointing out that there are other ways to think about it, too.

Let's throw one wrench into the mix before we get to our definition. Think about the "rule" that inputs a person in this class and outputs the color of their eyes. How would you write that down in the form $f(x) = \ldots$? It's hard! You'll essentially just have to rewrite the previous sentence in full as the definition of the "rule". What are the allowable inputs and outputs? They're not real numbers or integers or anything like that. They're something else entirely. Yet, this function is a perfectly reasonable one and we'd like to have it covered by our definition. Think about how this situation is (or isn't, really) different from something like $f(x) = x^2$ on $\mathbb{R}$. (You might even object here that this isn't a *function* at all! What about a person who has two different-colored eyes? What is the output of this "map" then? Oh boy!)

Okay, well, now it's your turn. Try to make a definition of a *function* using the concepts, terms, and mathematical objects we have discussed in these first few chapters.

7.2.1 Definition

Here's what we will use. Perhaps it's close to what your definition was, maybe they're identical, maybe our wordings are slightly different. Ultimately, though, this definition perfectly encapsulates the intuitive notion of a function that we had before (thinking of it as a *rule* of assignments), but it casts it in the language of sets and logic that we have been developing. This serves a few purposes: (a) it puts functions on a rigorous footing and allows us to confidently use them in a mathematical sense; (b) it allows us to discuss properties of functions and *prove* such things using mathematical terminology and concepts; and (c) it allows us to generalize the notion of a function and apply it in more abstract settings than just the standard sets of numbers we are familiar with. Alright, enough explanation, let's get to the definition.

Definition 7.2.1. *Let A, B be sets. Let f be a* relation *between A and B, so $f \subseteq A \times B$. Also, assume that f has the property that*

$$\forall a \in A \bullet \exists ! \, b \in B \bullet (a, b) \in f$$

(Recall that "$\exists !$" means "there exists a **unique** *...", i.e. "there is one and only one ...")*

Such a relation is called a **function** *from A to B.*

We call A the **domain** *of the function and B the* **codomain** *of the function.*

We write

$$f : A \to B$$

to mean f *is a function* **from** A **to** B.

If $(a, b) \in f$*, then we write*

$$f(a) = b$$

knowing that b *is the* unique *element that satisfies that property for the given* a.

That's it! It might be strange now to think about a *function* as a *relation*, which is actually a particular type of *set*, but that's what it is. Formulating functions in these terms allows us to talk about them in the language of sets and relations, but notice that we will still be able to use some familiar notation. Knowing that for every "input" a (i.e. every element of the *domain*), there is a *unique* "output" b (i.e. an element of the *codomain*), we can write $f(a) = b$ and know that the "=" is, actually, an equality. There is no other element of B that could satisfy this relationship, because that b is unique.

Part of this definition incorporates the idea we mentioned above: we want to know what *type* of object a function will "output". This is what specifying the codomain accomplishes. For example, it wouldn't make sense to define the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \sqrt{x}$; there are some elements (namely, negative numbers) of the domain where the "output" would be undefined. (Technically, the output would be a complex number, which is not an element of the codomain $\mathbb{R}$; in the context of $\mathbb{R}$, though, we would think of a complex number as being "undefined".) When a function is defined properly, and the domain and codomain are specified, and the related pairs do belong to the Cartesian product of the sets, we say the function is **well-defined**. Sometimes, we will present you with a relation on two sets and ask you to decide whether or not it is a *well-defined function*. In that case, we are really just asking whether the relation corresponds to a proper function.

The word "Range"

The word *codomain* might be new to you. In fact, you might be more accustomed to using the word **range** to refer to the set of potential "outputs" of a function. We want to completely avoid using the word "range" in this context because of a potential ambiguity. Some authors and teachers use the word "range" to mean what we mean here by "codomain": the set of *potential* "outputs" of a function. However, some authors and teachers use the word to mean what we mean here by "image". As you will see when we define this term properly in Section 7.3, this is the set of *actually-achieved* "outputs" of the function. In general, the image is a subset of the codomain, but it might be a *proper* subset. When someone uses the word "range", they might be thinking of one of these interpretations, but you might be thinking of the other one! To avoid this confusion, we will only use the words *codomain* and *image*.

7.2.2 Examples

Let's see some examples (and non-examples) of functions, using our new definition. While working with these examples, we will describe proper *notation* for defining functions and working with them, and we will describe how to "visualize" some functions and appeal to our intuitions.

Notation

There are several ways of properly defining a function. The following are all acceptable ways of defining "the squaring function on the real numbers":

> Let $f \subseteq \mathbb{R} \times \mathbb{R}$ be the function defined by $(x, y) \in f \iff y = x^2$.
>
> Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $f = \left\{(x, x^2) \mid x \in \mathbb{R}\right\}$.
>
> Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $\forall x \in \mathbb{R}\text{.}\ f(x) = x^2$.

Think about how each of these appeals to the definition of a *function* we saw above. The first appeals directly to the idea that a function is a type of *set*, namely a *relation* from $\mathbb{R}$ to $\mathbb{R}$. The second uses the same idea but expresses f via set-builder notation instead of an *if and only if* statement. The third appeals to the idea that every "input" of f has a *unique* "output", so we can simply declare what that "output" is *for all* $x \in \mathbb{R}$.

We will *usually* stick to the third notation style because it easier to understand, and appeals more directly to our intuitions about functions. Sometimes, we will use the other notation styles; we might be trying to emphasize the underlying structure of a function, or it might just be easier to write, depending on the context. In general, though, when defining a function, we need to specify all the important components for the reader: the *domain*, the *codomain*, the *letter name*, and either a *rule* or a *set* that assigns the ordered pairs.

If you're wondering why it's so important to specify the *codomain* when defining a function, think of it in terms of writing computer code. If you define a function, you usually have to *declare* the object type of the output variable. (This depends on the language, of course.) For instance, in `Java`, you might write

```
public int PlusOne (int x) {
    return x+1;
}
```

This defines a function that inputs an integer, adds one, and outputs another integer. Notice that you had to tell the program what type of object was *going in* and what type would be *coming out*.

Example 7.2.2. Consider the function that takes a natural number and outputs its binary representation. Let's use B to represent this function. Doing some calculating, we see that we want $B(1) = 1$ and $B(2) = 10$ and $B(10) = 1010$, for instance. What is the domain of this function? What is the codomain? Can

you write down the "rule" that defines it rigorously? Doesn't it seem better to just leave it the way we stated it, in words?

We would define this function in the following way. Let S be the set of all finite binary strings, i.e. sequences of 0s and 1s. Then we let $B : \mathbb{N} \to S$ be the function defined by

$$B = \{(n, s) \mid n \in \mathbb{N} \text{ and } s \text{ is the binary representation of } n\}$$

Example 7.2.3. Consider the "squaring function" again: Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $\forall x \in \mathbb{R}.\ f(x) = x^2$. Is this function any different from the following functions?

- Let $g : \mathbb{R} \to \mathbb{C}$ be the function defined by
$$\forall x \in \mathbb{R}.\ g(x) = x^2$$
- Let $h : \mathbb{Z} \to \mathbb{R}$ be the function defined by
$$\forall x \in \mathbb{Z}.\ h(x) = x^2$$

The function g has a different codomain but, in fact, $\mathbb{R} \subseteq \mathbb{C}$. All of the ordered pairs $(x, x^2) \in g$ still satisfy $x \in \mathbb{R}$ and $x^2 \in \mathbb{R}$. In this sense, f and g are the *same* function, and we would write $f = g$. We will see later on precisely what it means for two functions to be equal. For now, it suffices to say that the underlying relations corresponding to f and g have the same ordered pairs of real numbers as elements. The function g theoretically *allows* for complex numbers in the second coordinate, but the way the domain and the "rule" are established, this doesn't actually happen.

The function h has a different domain, and $\mathbb{Z} \subset \mathbb{R}$ (a *proper* subset). Thus, there are many ordered pairs in the relation corresponding to the function f that *don't* belong to the relation corresponding to the function h. For instance, $(1/2, 1/4) \in f$ but $(1/2, 1/4) \notin h$. Said another way, $f(1/2) = 1/4$, but the concept of $h(1/2)$ is not *well-defined*; $1/2$ does not belong to the domain of h.

Example 7.2.4. We can define a function **piece-wise**, as well. For instance, consider the *absolute value function*, defined on $\mathbb{R}$:

Let $a : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$\forall x \in \mathbb{R}. \quad a(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Every domain element falls into *exactly* one of the cases, so there is no ambiguity.

"Well-defined" Functions

It is not always clear that a defined relation *is* actually a function. Given a proposed domain, codomain, and a "rule" or set, how can we check that these represent a function? This is what the next definition (based entirely on the definition of function we saw above) addresses:

Definition 7.2.5. *Given a domain A, a codomain B, and a proposed "rule" for f, then we say f is a* **well-defined function** *if and only if (1) the rule is defined on* all *elements of A, and (2) for every $a \in A$, the rule outputs a* unique *element of the set B.*

Let's use this in an example here. Later in this section, we will see some *non-examples* of functions, and we will again appeal to this definition of **well-defined**.

Example 7.2.6. Let $f : \mathbb{Z} \to \mathbb{N}$ be the function defined by

$$\forall z \in \mathbb{Z}.\ f(z) = |2z + 1|$$

Wait, how can we be so sure that $|2z + 1|$ will be a *natural* number, for *any* integer z? It's not immediately obvious, but we can figure it out.

Suppose $z \in \mathbb{Z}$ satisfies $z \geq 1$. Then certainly $2z+1 \in \mathbb{Z}$ and $|2z+1| = 2z+1 \geq 3$. Thus, $f(z) \in \mathbb{N}$.

Suppose $z \in \mathbb{Z}$ satisfies $z \leq -1$. Then $2z + 1 \in \mathbb{Z}$ and $2z + 1 \leq -1$. Thus, $f(z) = |2z + 1| \geq 1$, so $f(z) \in \mathbb{N}$.

Suppose $z = 0$. Then $f(z) = |2 \cdot 0 + 1| = 1$, so $f(z) \in \mathbb{N}$.

In any case, we see that the "rule" that defines f does indeed yield a natural number, an element of the *codomain*. Furthermore, it yields *exactly* one such number. Therefore, this is a well-defined function.

Example 7.2.7. Let P be the set of all people in the world. Let $b : P \to \mathbb{N} \cup \{0\}$ be the function defined by

$$b = \{(p, n) \mid p \in P \wedge \text{ person } p \text{ has } n \text{ sisters}\}$$

(Notice that we have used one of the sets-emphasizing notation styles here, for practice. Also, it might look funny to combine math symbols and words, as in "$b(p)$ = the number of sisters person p has".)

Is this a well-defined function? We would say so. Walk up to someone (i.e. an element $p \in P$) and ask them how many sisters they have (i.e. what $b(p)$ is). They will tell you a non-negative integer in return. Also, they wouldn't possibly tell you two *different* numbers.

Now, you could point out that in today's society of divorces and remarriages, plenty of people have *half-sisters*, and $\frac{1}{2} \notin \mathbb{N} \cup \{0\}$. Fine. Fair point. But with the "simplifying assumption" that everyone has some *whole number* of sisters, this function is well-defined.

The Identity Function

Example 7.2.8. Let S be any set. *Must* there exist a function from S to S? Certainly, we can think of tons of functions from $\mathbb{R}$ to $\mathbb{R}$, but what if S is just some arbitrary set? Can we guarantee there is a function from S to S? It turns

out that ... yes, we can! Think back to a similar question we considered when talking about relations. (See Example 6.2.9.) We know that we can always define the *equality relation* on the set S; that is, we can define R on S by $(x, y) \in R \iff x = y$. This relation consists of all ordered pairs of the form (x, x), for every $x \in S$. Does this relation represent a function? We only need to check the definitional property: does every input have *exactly one* output? Sure looks like it! Any element of a set is only equal to itself, and nothing else. Thus, R does represent a function. This is a special enough function that we give it its own name.

Definition 7.2.9. *Let S be a set. The* **identity function** *on S is defined to be the function* $\mathrm{Id} : S \to S$ *given by*

$$\forall x \in S \bullet \mathrm{Id}(x) = x$$

That is, the identity function "outputs exactly what it inputs". (Thinking of a function as a machine, this is a lazy machine that does nothing, and just spits out whatever came in.)

Sometimes, we wish to refer to the identity functions as defined on *different* sets. To avoid confusion in that case, we will write Id_S to mean "the identity function **on the set** S".

Non-Examples

Sometimes, in the context of solving a problem, we might write down a proposed "rule" between two sets and wonder whether it is a function. Perhaps we need it to be a function to help us work something out. How could this *fail*? That is, what could we be looking for to show that a proposed rule is *not* a function? Look back at the definition of **well-defined function** (see Definition 7.2.5). Three different things could go wrong:

- There might be an element of the domain for which **no "output"** is defined.
- There might be an element of the domain for which **more than one "output"** is defined.
- There might be an element of the domain for which exactly one "output" is defined, but it is **not** an element of the **codomain**.

The following examples illustrate these possibilities.

Example 7.2.10. Let $G : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be defined by

$$\forall (a, b) \in \mathbb{N} \times \mathbb{N} \bullet G(a, b) = a - b$$

This is **not** a well-defined function because there are many domain elements whose "output" is **not** an element of the codomain. For instance, $(5, 10) \in \mathbb{N} \times \mathbb{N}$ and $G(5, 10) = -5$, but $-5 \notin \mathbb{N}$.

Example 7.2.11. Let W be the set of words in the English language. Let $A : W \to W$ be defined by taking in a word and outputting an anagram of that word that is not the original word. This is **not well-defined** for several reasons. For example, A(HI) does not exist; the same goes for A(FUNCTION). Also, notice that, for instance, A(INTEGRAL) has multiple (i.e. non-unique) outputs: TRIANGLE, ALERTING, ALTERING, ...

7.2.3 Equality of Functions

It is sometimes the case that two functions are defined by different "rules" or formulas, but they correspond to the same underlying relation! In that sense, the two functions are **equal**. We would like to describe what this means in terms of function notation, first, and then we will prove this idea we've come up with using the underlying relation and set notation.

How might two functions be *equal*? Certainly their domains must be the same. If not, then one of the domain sets contains an element that doesn't belong to the other domain set, and this is a problem. (Think about it: one of the functions would be defined on an element for which the other function is *undefined*, so there's no way they can be equal). So, let's say we have two sets, A and B, and two functions, $f : A \to B$ and $g : A \to B$. What do we require of f and g for them to be equal? The defining characteristic of the functions is that for any input, say $x \in A$, there is a *unique* output $f(x)$ and a *unique* output $g(x)$. If f and g are to be the same function, we better have $f(x) = g(x)$! This lets us state the following theorem.

Theorem 7.2.12. *Let A, B be sets, and let $f : A \to B$ and $g : A \to B$ be functions. Suppose that*

$$\forall x \in A \bullet f(x) = g(x)$$

Then we say f and g are **equal** *as functions, and we write $f = g$.*

This is indeed a *theorem*. The idea is very intuitive, but it is not explicitly part of the *definition* of a function. Thus, we have to *prove* this idea. To complete this proof, we will consider the ordered pairs that belong to the relations f and g. By showing that these ordered pairs are the *same*, we can conclude $f = g$ in the sense of *sets*. Notice that we are making a *double-containment* argument!

Proof. Let A, B be sets, and let $f : A \to B$ and $g : A \to B$ be functions. Suppose that

$$\forall x \in A \bullet f(x) = g(x)$$

First, we prove $f \subseteq g$. Let $(a, b) \in f$ be given. Since f is a function, this means $f(a) = b$. By the main assumption, $f(a) = g(a)$, and so $g(a) = b$, as well. Thus, $(a, b) \in g$. This shows that

$$(a, b) \in f \implies (a, b) \in g$$

and, therefore, $f \subseteq g$.

Second, we prove $g \subseteq f$ in a similar manner. Let $(c, d) \in g$ be given. Since g is a function, this means $g(c) = d$. By the main assumption, $f(c) = g(c)$, and so, $f(c) = d$, as well. Thus, $(c, d) \in f$. This shows that

$$(c, d) \in g \implies (c, d) \in f$$

and, therefore, $g \subseteq f$.

Since we have shown $f \subseteq g$ and $g \subseteq f$, we may conclude $f = g$. □

This provides us with a handy way of showing that two functions are equal without having to delve into the underlying relation/set structure. Instead, we just have to show that every element of the domain produces the same output under both functions. Let's see how this works in a couple of examples.

Example 7.2.13. Let $A = \{-1, 0, 1\}$. Define the functions $f_1 : A \to \mathbb{Z}$ and $f_2 : A \to \mathbb{Z}$ by

$$\forall x \in A \bullet f_1(x) = x \ \wedge \ f_2(x) = x^3$$

Let's prove that $f_1 = f_2$. Since the domain only contains three elements, we can verify these outputs one by one. Notice that

$$\begin{aligned} f_1(-1) = -1 = (-1)^3 = f_2(-1) \\ f_1(0) = 0 = 0^3 = f_2(0) \\ f_1(1) = 1 = 1^3 = f_2(1) \end{aligned}$$

Thus, for every allowable input (i.e. $\forall x \in A$) the functions f_1 and f_2 have the same output (i.e. $f_1(x) = f_2(x)$). Therefore, $f_1 = f_2$.

Example 7.2.14. Let $B = \{1, 2, 3\}$. Define the functions $g_1 : B \to \mathbb{Z}$ and $g_2 : B \to \mathbb{Z}$ by

$$\forall n \in B \bullet g_1(n) = n^3 - n^2 - 6 \ \wedge \ g_2(n) = 5n^2 - 11n$$

Let's prove that $g_1 = g_2$. Again, we only have three elements to consider, so we can just verify all of the equalities by hand:

$$\begin{aligned} g_1(1) &= 1^3 - 1^2 - 6 = 1 - 1 - 6 = -6 \\ g_2(1) &= 5 \cdot 1^2 - 11 \cdot 1 = 5 - 11 = -6 \end{aligned}$$

$$\begin{aligned} g_1(2) &= 2^3 - 2^2 - 6 = 8 - 4 - 6 = -2 \\ g_2(2) &= 5 \cdot 2^2 - 11 \cdot 2 = 20 - 22 = -2 \end{aligned}$$

$$\begin{aligned} g_1(3) &= 3^3 - 3^2 - 6 = 27 - 9 - 6 = 12 \\ g_2(3) &= 5 \cdot 3^2 - 11 \cdot 3 = 45 - 33 = 12 \end{aligned}$$

Thus, $\forall n \in B \bullet g_1(n) = g_2(n)$, and so $g_1 = g_2$.

Since the domains in these two examples were "small", we were able to examine every element one-by-one. This is not always the case, though. Sometimes, we must consider an *arbitrary* element of the domain (since the desired property we are proving begins with a "$\forall$" quantifier) and work with it. As it turns out, there is an interesting way of doing that with this example, so let's show you that now to get an idea of how it works.

Let $n \in B$ be given. Since $g_1(n), g_2(n) \in \mathbb{Z}$, we can consider their difference. Specifically, we see that

$$\begin{aligned} g_1(n) - g_2(n) &= (n^3 - n^2 - 6) - (5n^2 - 11n) \\ &= n^3 - 6n^2 - 11n - 6 \\ &= (n-1)(n-2)(n-3) \end{aligned}$$

(Note: the reader can verify the last equality by simply expanding the product of the three terms.)

Since $n \in B$, we know $n = 1$ or $n = 2$ or $n = 3$. In each case, one of the terms—either $n-1$ or $n-2$ or $n-3$—must be zero. Thus,

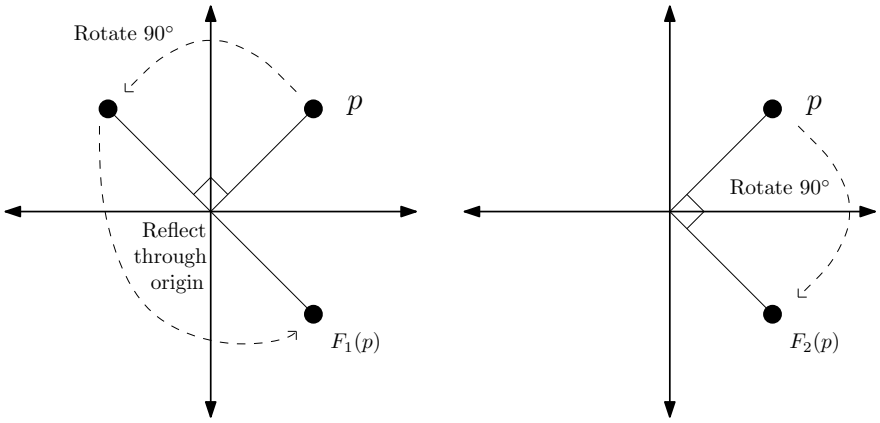
$$g_1(n) - g_2(n) = (n-1)(n-2)(n-3) = 0$$

Accordingly, by adding to both sides, we find $g_1(n) = g_2(n)$. Since this holds for arbitrary $n \in B$, we conclude that $g_1 = g_2$.

We will remark that this is certainly trickier than the "check all the cases" approach in this specific example. How did we know to look at the difference? How did we know that it would factor like that?! This is why mathematics is so interesting! We might have the ingenuity to check something like that by thinking about how to approach a more general problem, where the domain is too large to consider every case one by one. We might recognize the factorization, or think to guess at it from the fact that $B = \{1, 2, 3\}$. And if you think about it carefully enough, you might realize how we came up with this example! ☺

Let's look at one final example of equality of functions which is a bit more complicated. It involves some ideas that we haven't assumed any familiarity with, and we won't discuss them again, but we found it interesting and illustrative enough to include it here.

Example 7.2.15. We will define two functions here and argue why they are equal. Let the domain and codomain of each function be $\mathbb{R}^2$, the real plane. Now, let's define two functions, F_1 and F_2 by describing their actions geometrically (i.e. visually). We want F_1 to input a point on the plane—let's call it p—and output the point achieved by rotating p counterclockwise by $90°$ around the origin and then reflecting through the origin. We want F_2 to input p and output the point achieved by rotating p clockwise by $90°$. To get a better idea of what this means, look at the following pictorial representations of F_1 and F_2 applied to a particular point.



We claim that $F_1 = F_2$, in the sense of functions. By playing around with several examples, we can see that this *might* be true; that is, we can't come up with counterexamples, and we might even begin to "see" why it's true. But none of this is a rigorous proof. It's just an intuitive way of understanding something. To rigorously prove this fact, we'll need to use some mathematical machinery outside the scope of this class. For that reason, we'll really only "sketch" this proof, and leave some of the technical details to be explored by interested readers.

The main idea is this: we can represent points in the plane by *vectors*, the functions by *matrices*, and the action of a function on a point by matrix multiplication. Do not worry if you have no knowledge of matrices or vectors; you can just skip this example and you won't be missing anything essential! If you'd like to follow along, though, we'll say this: matrices are just arrays of real numbers, and vectors are matrices with just one column. The point $(1, 1)$ in the real plane is represented by the vector $\begin{bmatrix}1\\1\end{bmatrix}$, for instance. The action of *rotation* of a vector can be represented by a *rotation matrix*. (You might see this in an intermediate physics course, like electromagnetics or mechanics.) For example, the action of rotation counterclockwise by 90° can be represented by the matrix

$$\begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$$

Rotating the point (a, b) counterclockwise by 90° amounts to multiplying the corresponding vector by this matrix, following the usual matrix multiplication rules (where we multiply a row on the left by a column on the right, entry by entry, and add). As an example, let's rotate the point $(1, 1)$:

$$\begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0+(-1)\\1+0\end{bmatrix} = \begin{bmatrix}-1\\1\end{bmatrix}$$

Look at that, it matches up with what we'd expect! Check out the picture above to see this rotation in action.

Similarly, we can represent clockwise rotation by 90° by this matrix:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

(Note the similarities in the entries, even. There's a reason for that, which we'll leave you to discover via some Googling. Or Binging, we suppose.)

Reflection through the origin can also be represented by matrix multiplication, but there's an even easier way to think about it: just negate both of the coordinates. For instance, the reflection of $(-1, 1)$ through the origin is $(1, -1)$.

This allows us to fully represent the actions of F_1 and F_2. Since F_1 says "rotate counterclockwise by 90° and negate both entries", we can write

$$F_1\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = -\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} b \\ -a \end{bmatrix}$$

(where the $-$ sign out front accomplishes the negation), and since F_2 says "rotate clockwise by 90°", we can write

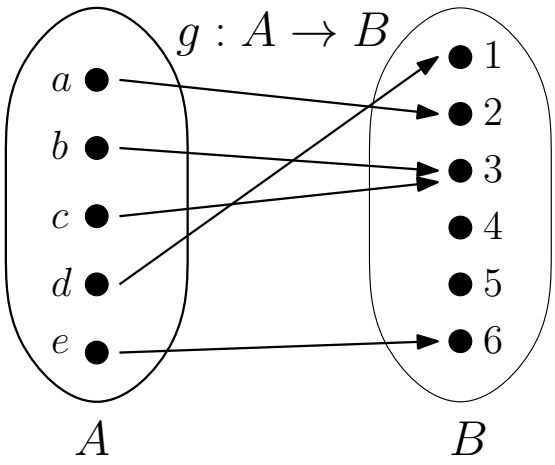
$$F_2\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} b \\ -a \end{bmatrix}$$

By following the rules of matrix and vector multiplication, we can easily see that these two expressions are equal, for any a and b. So, assuming some knowledge about rotation matrices, this proves $F_1 = F_2$!

7.2.4 Schematics

Before we move on to talking about some more abstract properties of functions and how to prove them, we will describe one helpful way of representing functions. We want to emphasize that this method is not *rigorous* in a mathematical sense, and using these representations in a *proof* is probably not the best idea. (On a graded homework problem, say, you might not receive full credit, despite "having the right idea".) However, this method does provide some intuitive insight into how functions work, and can guide you into discovering something and figuring out how to prove it more rigorously. In particular, this method will be quite helpful in constructing *counterexamples* to particular claims about function properties.

The idea of a **schematic diagram** is similar to how we used *venn diagrams* to represent sets. A set is a collection of elements, not a shaded circle on a piece of paper, but these shaded circles and their overlaps can help us figure out and describe something about sets. Likewise, a function is a relation on two sets with a particular property, not something like this:



However, this does somehow *represent* the idea of a function. In this picture, we have represented the domain A by an oval, and the same with the codomain B. The elements of A and B are represented by dots inside those ovals (and they are labeled), and we have drawn arrows between those dots based on what the function $g : A \to B$ does.

Mostly, this method is used to explore a certain property of a function and perhaps construct a counterexample to a claim. By drawing some dots and arrows and playing around with how they connect, we can perhaps develop the underlying *structure* of an example; then, we can go back and assign some names and formulas to the parts of the diagram and make the picture more rigorous.

We will use some schematic diagrams to illustrate some properties and concepts as we proceed, but these will always be accompanied by a more rigorous statement or description. We encourage you to employ a similar method.

7.2.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) Write down the definition of a **function**, *without* looking it up. Then, compare to our definition. Does yours convey the same information? If not, what did you miss?

(2) What is the difference between the **domain** and the **codomain** of a function?

(3) What does it mean for a function to be **well-defined**?

(4) What is the **identity function** and how is it defined?

(5) How can we prove that two functions are **equal**?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Use proper notation to define a function that inputs an integer and outputs the square root of its absolute value.

What is the domain of this function? What is its codomain?

(2) Use proper notation to define a function that inputs a pair of natural numbers and outputs their average (arithmetic mean).

What is the domain of this function? What is its codomain?

(3) Let $A = \{-2, -1, 0, 1, 2\}$. Let $g : A \to A$ be defined by $\forall x \in A.\ g(x) = x^2 - 3$. Draw a schematic diagram to determine whether g is well-defined or not. Is it?

(4) Let X be any set. Use proper notation to define a function that inputs a *subset* of X and outputs that set's complement (in the context of X).

What is the domain of this function? What is its codomain?

(5) Let $B = \{-1, 0, 1\}$. Let $h : B \to B$ be defined by $\forall b \in B.\ h(b) = b^3$. What special function is this equal to?

(6) Let $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{N}$ be defined by $\forall (x, y) \in \mathbb{Z} \times \mathbb{Z}.\ f(x, y) = \frac{1}{2}|x + 1| \cdot |y|$. Is this a well-defined function? Why or why not?

7.3 Images and Pre-images

7.3.1 Image: Definition and Examples

Think back to the definition of a function. We required that every input have a unique output. This ensures that a function is defined *everywhere* on its domain. What about the codomain, though? All we required there was that all of the *outputs* belong to the codomain. We never said anything about "how much" of the codomain is "covered". The idea of the **image** of a function is to capture exactly this notion. As we will see from some examples, it is not always easy to determine precisely what the image of a function is, even when we know what the codomain is. It is for this reason that we defined a *function* and its *codomain* first, before introducing the *image*; so don't think we were trying to fool you or anything!

Definition

Definition 7.3.1. *Let A, B be sets and let $f : A \to B$ be a function. Let $X \subseteq A$.*

The **image of X under the function f** *is written and defined as*

$$\mathrm{Im}_f(X) = \{b \in B \mid \exists a \in X \bullet f(a) = b\}$$

That is, the image of X under f is the set of all "outputs" that come from "inputs" in the set X.

An equivalent way of writing this is

$$\mathrm{Im}_f(X) = \{f(a) \mid a \in X\}$$

(We will sometimes abbreviate the notation as just $\mathrm{Im}(X)$, when the function is clearly identified and unambiguous, and consequently refer to the set as just "the image of X", instead of "the image of X under f".)

When we say **the image of f**, *we mean the image of the entire domain, i.e. $\mathrm{Im}_f(A)$.*

Notice that this is defined for *any* subset of the domain, $X \subseteq A$, so we can talk about the image of any "piece" of the domain, or all of it. We will see some examples now—and exercises later—that consider strict subsets $X \subset A$, as well as A itself.

One Observation

Notice that

$$\mathrm{Im}_f(A) \subseteq B$$

no matter what f and A and B are. This follows *by definition*, since we used set-builder notation to define the image via elements of B. In the next section, we will explore what happens when $\mathrm{Im}_f(A) = B$.

For now, let's practice identifying the images of certain functions. In some cases, we will be provided with a function and its image and asked to verify this claim, but in other cases, we will need to develop some techniques to figure out what the image is in the first place!

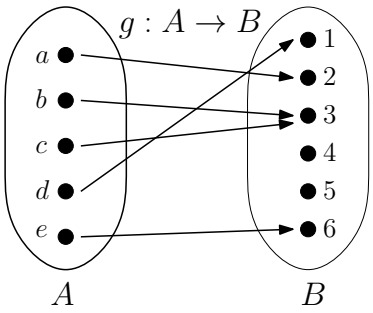
Examples

Example 7.3.2. Define a function $g : A \to B$ by setting $A = \{a, b, c, d, e\}$ and $B = \{1, 2, 3, 4, 5, 6\}$ and

$$g = \{(a, 2), (b, 3)(c, 3), (d, 1), (e, 6)\}$$

Define $X_1 = \{a, b, c\}$ and $X_2 = \{a, c, e\}$ and $X_3 = \{c, d, e\}$.

You might notice that this is the same function we defined in the schematic diagram in the last section! Let's see that diagram again, because it can help us identify the images in the following list.



(1) $\text{Im}_g(\{a\}) = \{2\}$

This is because $g(a) = 2$.

Notice the use of *set brackets*. We always find the image of a *set*, so writing $\text{Im}_g(a)$ would be *incorrect*.

(2) $\text{Im}_g(\{b, c\}) = \{3\}$

This is because $g(b) = g(c) = 3$.

(3) $\text{Im}_g(X_1) = \{2, 3\}$

This is because $g(b) = g(c) = 3$ and $g(a) = 2$.

(4) $\text{Im}_g(X_2) = \{2, 3, 6\}$

This is because $g(a) = 2$ and $g(c) = 3$ and $g(e) = 6$.

(5) $\text{Im}_g(X_3) = \{1, 3, 6\}$

This is because $g(c) = 3$ and $g(d) = 1$ and $g(e) = 6$.

(6) $\text{Im}_g(A) = \{1, 2, 3, 6\}$

Looking at the set B in the schematic diagram, we see that these are the only values that are "hit" by the function. Notice $4, 5 \in B$ but $4, 5 \notin \text{Im}_g(A)$, so $\text{Im}_g(A) \subset B$ (a *proper* subset).

Example 7.3.3. Consider the temperatures (in degrees Celsius) where water is in its liquid state. Specifically, define the set

$$C = \{x \in \mathbb{R} \mid 0 < x < 100\}$$

and define the function $F : C \to \mathbb{R}$ by

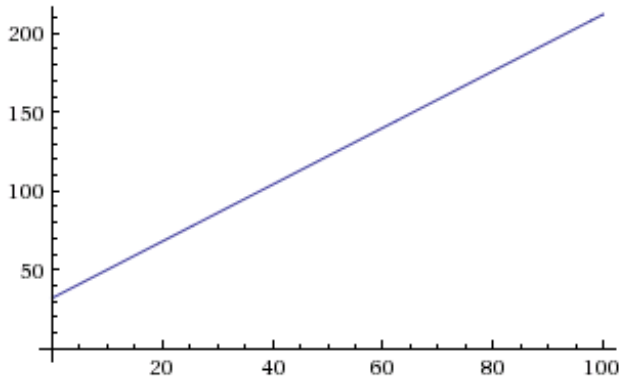
$$\forall c \in C\,.\; F(c) = \frac{9}{5}c + 32$$

What is $\text{Im}_F(C)$? What does it represent?

To approach questions like these, we must (a) identify a *claim* for what $\text{Im}_F(C)$ is by defining a set, and then (b) prove that the set we defined is actually *equal*

to $\text{Im}_F(C)$, in the sense of sets. This means we will use a *double-containment argument*!

Solution: Define $S = \{y \in \mathbb{R} \mid 32 < y < 212\}$. (Notice that this represent the set of temperatures (in degrees Fahrenheit) where water is in its liquid state.) We claim $S = \text{Im}_F(C)$.



It is hard to give advice about how to *come up with* claims like this, in general. Most often, this relies on some playing around with the function and testing some values, and perhaps some insight about some other properties of the function. In this specific case, we notice that this function is *increasing*; that is, if we have two input values with $c_1 < c_2$, then we know that $F(c_1) < F(c_2)$. We can glean this information from the graph of the function (see above) and/or recognizing it is a *linear* polynomial. Accordingly, to identify the image, we just have to consider the smallest and largest inputs and identify their outputs. (Again, we can glean this information from a graph.) We find that

$$F(0) = 0 + 32 = 32 \quad \text{and} \quad F(100) = \frac{900}{5} + 32 = 212$$

From this, we defined the set S. (Also, notice that we had to use "$<$" in the inequality because, in fact, $0 \notin C$, the domain!) We also give this set a name so that we can refer to it later without implicitly claiming, already, that it *is* the image. This is a somewhat subtle distinction, but an important one! Now, let's prove our claim.

Proof. First, we'll prove $\text{Im}_F(C) \subseteq S$. (In other words, we'll prove that every output of the function F actually satisfies the inequality in the definition of S.)

(To do this we will start with an arbitrary element of $\text{Im}_F(C)$, and appeal to the *defintion* of image to bring an element of the *domain* into play.)

Let $y \in \text{Im}_F(C)$ be arbitrary and fixed. By the definition of image, this means $\exists x \in C$ such that $F(x) = y$. Let such an x be given.

By the definition of C, we know $0 < x < 100$. By the definition of F, we know

$F(x) = \frac{9}{5}x + 32$. Since multiplying by a *positive* number and adding to both sides *preserves* inequalities, we can deduce that

$$\frac{9}{5} \cdot 0 + 32 < F(x) < \frac{9}{5} \cdot 100 + 32$$

and, simplifying, this tells u

$$32 < F(x) < 212$$

Thus, $F(x) \in S$, i.e $y \in S$. Therefore, $\text{Im}_F(C) \subseteq S$.

Second, we'll prove $S \subseteq \text{Im}_F(C)$. (In other words, we'll prove that every element of S is actually "achieved" by the function F somehow. This amounts to proving an existential claim, i.e. that some element of the domain *exists*.)

Let $s \in S$ be arbitrary and fixed. By the definition of S, we know that $s \in \mathbb{R}$ and $32 < s < 212$. We need to prove that $\exists c \in C \,.\, F(c) = s$.

(By doing some scratch work on the side, that you can work through on your own, we came up with the following idea. Just set an expression equal to s and solve for c ...)

Define $c = \frac{5}{9}(s - 32)$.

Let's show $c \in C$. By using the information we have about s and manipulating the inequalities, we observe that

$$\begin{aligned} 32 < s < 212 &\implies 0 < s - 32 < 180 \\ &\implies 0 < \frac{5}{9}(s - 32) < \frac{5}{9} \cdot 180 = 100 \\ &\implies 0 < c < 100 \end{aligned}$$

Since $c \in \mathbb{R}$, certainly, this shows that $c \in C$.

Next, let's show that $F(c) = s$. We observe that

$$\begin{aligned} F(c) = \frac{9}{5}c + 32 &= \frac{9}{5}\left(\frac{5}{9}(s - 32)\right) + 32 \\ &= (s - 32) + 32 = s \end{aligned}$$

Together, this shows that $s \in \text{Im}_F(C)$, as well. Thus, $S \subseteq \text{Im}_F(C)$.

Overall, by a double-containment argument, we conclude that $S = \text{Im}_F(C)$. $\square$

The second half of our proof is certainly the harder part, and this is generally true of proofs like this. In coming up with a candidate c, we essentially have to "undo" the process that the function F does and find an input c for our given output s. In a case like this, where the function is a numerical/arithmetical operation on real numbers, the best route is to set up the desired equality, like

$$\frac{9}{5}c + 32 = s$$

and solve the equality for c. This function is linear, so this process only produces one such s but, in general, we might expect multiple values of s to work. Ultimately, we only need *one* working value to complete this part of the proof, so we can just select *any* one that works and use that as our claim. Sometimes, though, this makes it harder to find such a value. It all depends on the example at hand. Other times, we might be working with functions that aren't defined on sets of numbers, and we have to use some more abstract insight to come up with a candidate element. Again, this all depends on the given situation, and with practice, you'll become much better at it!

Oh right, we asked what this image represents! Since the domain represented the temperatures, in degrees Celsius, at which water is a liquid, the image represents the temperatures, in degrees Fahrenheit, at which water is a liquid.

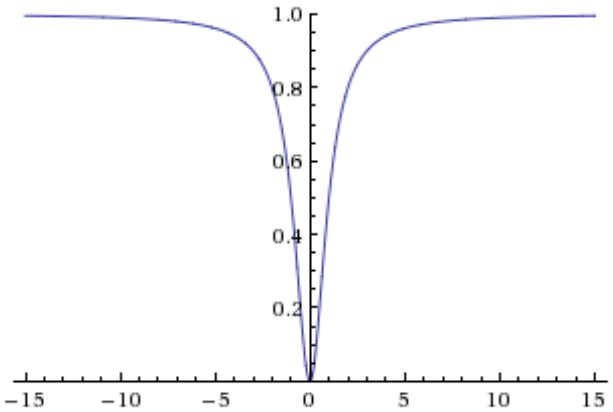
Let's look at another example of proving the image of a function is a particular set.

Example 7.3.4. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$\forall x \in \mathbb{R}.\ f(x) = \frac{x^2}{1 + x^2}$$

Let's determine the image, $\mathrm{Im}_f(\mathbb{R})$, and prove our claim!

Here, again, we must use some outside strategies and intuition to identify the image first. Using some techniques from calculus or algebra, we could plot the graph of this function and try to guess the image. Try that if you'd like. You'll end up with this graph:



We can also recognize that the denominator is greater than the numerator and so, as x gets larger and larger, those two quantities get closer and closer together, relatively speaking. (That is, their ratio approaches 1.) Also, both terms are nonegative, since they involve squares, so their ratio is at least 0. In any event, we can piece our observations together and make the following claim:

Define the set

$$T = \{y \in \mathbb{R} \mid 0 \leq y < 1\}$$

We claim that $T = \text{Im}_f(\mathbb{R})$.

We now follow the same type of strategy we employed in the previous example. Before we do so, let's remember that the second part of that proof—showing the claimed set is a subset of the image—was the harder part, and try to anticipate what will happen there.

In that part, we will be working with an arbitrary element $y \in T$ and we will want to find an element $x \in \mathbb{R}$ that satisfies $f(x) = y$. To find such an x, let's set up the equality and try to solve for x:

$$\begin{aligned} y = \frac{x^2}{1+x^2} &\iff (1+x^2)y = x^2 \\ &\iff y + yx^2 - x^2 = 0 \\ &\iff (y-1)x^2 = -y \\ &\iff x^2 = \frac{y}{1-y} \end{aligned}$$

Now what? Can we be assured $\frac{y}{1-y} \in \mathbb{R}$, even? Can we be assured it's nonnegative, so that there exists such an x? What about the fact that there might be *two* roots? Think about these potential issues and try to write your own version of this proof before reading on for ours!

Proof. First, let's prove $\text{Im}_f(\mathbb{R}) \subseteq T$.

Let $y \in \text{Im}_f(\mathbb{R})$ be arbitrary and fixed. By the definition of image, we know $\exists x \in \mathbb{R}$ such that $f(x) = y$. Let such an x be given.

Since $x \in \mathbb{R}$, we know $x^2 \geq 0$ and $0 < x^2 + 1$. We can then deduce that $0 < \frac{1}{x^2+1}$.

By multiplying the previous two inequalities—which we can do since all the terms are non-negative—we may deduce that $0 \leq \frac{x^2}{1+x^2}$.

Next, we know that $0 \leq x^2 < x^2 + 1$, so $\frac{x^2}{1+x^2} < 1$, as well. (Note: why was it important to point out that $x^2 \geq 0$? What can go wrong there?)

This shows that $0 \leq \frac{x^2}{1+x^2} < 1$. Since $y = f(x) = \frac{x^2}{1+x^2}$, this is equivalent to saying $0 \leq y < 1$.

Thus, $y \in T$, and so $\text{Im}_f(\mathbb{R}) \subseteq T$.

Second, let's prove $T \subseteq \text{Im}_f(\mathbb{R})$.

Let $y \in T$ be arbitrary and fixed. This means $y \in \mathbb{R}$ and $0 \leq y < 1$.

To show $y \in \text{Im}_f(\mathbb{R})$, as well, we must produce an x such that $f(x) = y$.

We claim that $x = \sqrt{\frac{y}{1-y}}$ works.

Notice that $y \geq 0$, and $y < 1$ implies $-y > -1$ so $1 - y > 0$. Thus, $\frac{y}{1-y} \geq 0$, and so $x \in \mathbb{R}$ is well-defined as a square root, and x belongs to the domain $\mathbb{R}$.

Next, notice that $x^2 = \frac{y}{1-y}$, and so

$$f(x) = \frac{x^2}{1+x^2} = \frac{\frac{y}{1-y}}{1+\frac{y}{1-y}} = \frac{\frac{y}{1-y}}{\frac{(1-y)+y}{1-y}} = \frac{\frac{y}{1-y}}{\frac{1}{1-y}} = \frac{y}{1-y} \cdot \frac{1-y}{1} = \frac{y}{1} = y$$

This shows that $y \in \text{Im}_f(\mathbb{R})$, and so $T \subseteq \text{Im}_f(\mathbb{R})$.

Overall, by a double-containment proof, we conclude that $T = \text{Im}_f(\mathbb{R})$. □

Notice how we addressed the issues discussed before the proof. Yes, two potential x values existed that would work (namely, the + and − square roots) but we only *needed* one, so we just picked one (the positive one) and ran with it.

(Questions: What if this function was defined only on the nonnegative real numbers? What about just the negative real numbers? How might that restriction affect our choice?)

Example 7.3.5. Consider the function $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by

$$\forall (a, b) \in \mathbb{N} \times \mathbb{N}\text{.}\ p(a, b) = ab + a$$

What is $\text{Im}_p(\mathbb{N} \times \mathbb{N})$?

This example might feel a little trickier because the domain is a Cartesian product of sets; that is, p inputs an ordered pair of natural numbers and outputs a single natural number. A good approach in a situation like this is to just start plugging in some values and seeing what happens. Consider the following table of values as a way to get started, where the left column indicates values of a, the top row indicates values of b, and the table entries are the values of $p(a, b)$.

	1	2	3	4	5
1	2	3	4	5	6
2	4	6	8	10	12
3	6	9	12	15	18
4	8	12	16	20	24
5	10	15	20	25	30

It looks like every natural number is "achieved" by the function p, except for 1. Specifically, look at the top row of the array of values: there are all the natural numbers except 1. Let's use this insight in the following proof.

Proof. Let $V = \mathbb{N} - \{1\}$. We claim $V = \text{Im}_p(\mathbb{N} \times \mathbb{N})$.

First, we prove $\text{Im}_p(\mathbb{N} \times \mathbb{N}) \subseteq V$. Let $n \in \text{Im}_p(\mathbb{N} \times \mathbb{N})$ be arbitrary and fixed.

This means $n \in \mathbb{N}$ and $\exists (a, b) \in \mathbb{N} \times \mathbb{N}$ such that $p(a, b) = n$. Let such (a, b) be

given.

This means $n = ab + a$. Since $a, b \geq 1$, then $ab \geq 1$ and so $n = ab + a \geq 2$. By the defintion of V, this shows that $n \in V$.

Thus, $\text{Im}_p(\mathbb{N} \times \mathbb{N}) \subseteq V$.

(Try to write the next half of the proof before reading on and seeing ours!)

Second, we prove $V \subseteq \text{Im}_p(\mathbb{N} \times \mathbb{N})$. Let $v \in V$ be arbitary and fixed.

This means $v \in \mathbb{N}$ and $v \geq 2$. Define $(a, b) = (v - 1, 1)$.

Notice that $v - 1 \geq 1$, so $v - 1 \in \mathbb{N}$ and thus $(a, b) \in \mathbb{N} \times \mathbb{N}$.

Also, notice that

$$p(a, b) = p(v - 1, 1) = (v - 1) \cdot 1 + 1 = v - 1 + 1 = v$$

Thus, $p(a, b) = v$, and so $(a, b) \in \text{Im}_p(N \times \mathbb{N})$. Therefore, $V \subseteq \text{Im}_p(\mathbb{N} \times \mathbb{N})$.

By a double-containment proof, we have shown $V = \text{Im}_p(\mathbb{N} \times \mathbb{N})$. □

7.3.2 Proofs about Images

You might have observed the following fact by playing around with some of the examples we have seen. Either way, we can make state and prove this claim by working with the definition of image. Notice that it is a claim about an *arbitrary* function; it holds no matter what f is!

Proposition 7.3.6. *Let A, B be sets. Let $f : A \to B$ be a function. Let $S, T \subseteq A$. Then,*

$$\text{Im}_f(S \cap T) \subseteq \text{Im}_f(S) \cap \text{Im}_f(T)$$

Proof. Let $z \in \text{Im}_f(S \cap T)$ be arbitrary and fixed. This means $\exists a \in S \cap T$ such that $f(a) = z$. Let such an a be given.

Since $a \in S \cap T$, we know $a \in S$ and $a \in T$.

Thus, $z \in \text{Im}_f(S)$ and $z \in \text{Im}_f(T)$, by the definition of image.

We deduce that $z \in \text{Im}_f(S) \cap \text{Im}_f(T)$, by the definition of intersection.

This shows the desired set containment. □

Why didn't we claim an *equality* here? It turns out that equality *need not hold*, in fact! That is, there exists at least one function such that the reverse containment—namely, $\text{Im}_f(S) \cap \text{Im}_f(T) \subseteq \text{Im}_f(S \cap T)$—is False. We will provide an example of such a function below.

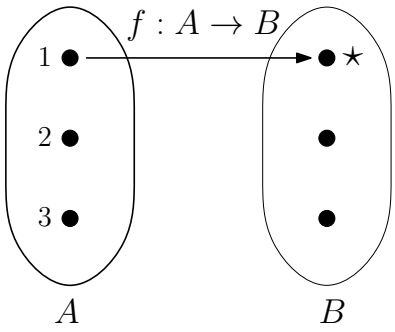
(You should try to come up with an example of a function where this reverse containment *does* hold. Together, we will have shown that one cannot make a conclusion that *necessarily* holds about this containment!)

We will use a schematic diagram to *come up* with an example with the desired properties. We will then use this to formally *define* a function and state its properties, pointing out how they match what will be established in our claim.

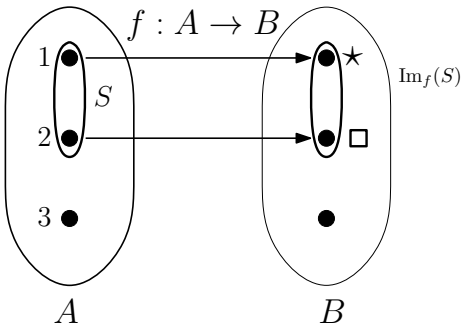
We want to point out that employing this technique is perfectly valid, as long as you go back and write down a formal definition afterwards. Turning in just a schematic diagram as a "proof" is not rigorous enough, but this can certainly help guide your intuition into producing fruitful *ideas* for a proof!

Furthermore, keep in mind that there is no need to construct the most *complicated* or *interesting* counterexample in situations like this. If you're trying to *disprove* a universally-quantified statement, you just need *one* example that works! In particular, don't feel like you need to define a function that works with *numbers*, using some *formula*. Sometimes, this will actually make your job much harder! It's typically the case that a counterexample can be made using sets with just a few (i.e. two or three) elements each.

Example 7.3.7. We claim that there exist sets A, B, S, T and a function $f : A \to B$ such that $\text{Im}_f(S) \cap \text{Im}_f(T) \not\subseteq \text{Im}_f(S \cap T)$. Let's figure out how to construct such an example. Based on our comment above, we are going to try to make an example where the sets involve three or so elements. Let's get the process started by taking A to be $\{1, 2, 3\}$ and defining $f(1)$:



Now, just to have a defintion in hand, let's choose $S = \{1, 2\}$. It seems like it will be more reasonable to work with 2 elements in S, so we'll make that choice. Also, it seems like we should make $f(1) \neq f(2)$. Otherwise, $\text{Im}_f(S)$ would contain only one element, and there would have been no point in making S have two elements. So let's define $f(2)$, as well:



Now, we need to choose T. It will be interesting to have $S \cap T \neq \varnothing$, but it would be hard to handle (perhaps) if $T \supseteq S$. So, let's say $T = \{2, 3\}$. Then, we just need to choose $f(3)$. In considering each of these cases, look at the schematic diagram above, and imagine drawing an arrow to represent $f(3)$.

- What if $f(3) = f(2) = \square$?

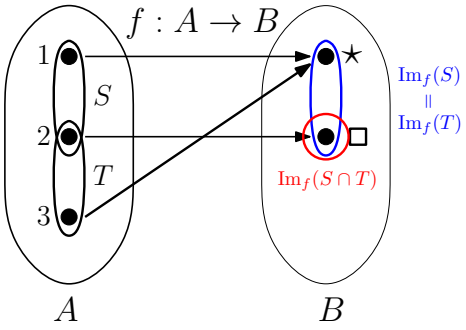
  In this case, $\mathrm{Im}_(T) = \{\square\}$, so $\mathrm{Im}_f(S) \cap \mathrm{Im}_f(T) = \{\square\}$. But $\mathrm{Im}_f(S \cap T) = \{\square\}$, as well! This doesn't work.

- What if $f(3)$ is something else, like $f(3) = \odot$?

  This doesn't work either! We will have $\mathrm{Im}_f(S) \cap \mathrm{Im}_f(T) = \{\square\} = \mathrm{Im}_f(S \cap T)$.

- What if $f(3) = f(1) = \star$?

  It looks like this works!



We have made it so that $\mathrm{Im}_f(S) \cap \mathrm{Im}_f(T)$ is a *strict* superset of $\mathrm{Im}_f(S \cap T)$.

Look back over our construction, and see if you understand our thought process. What were the restrictions we had to conform to? Where did we have freedom of choice? What did we decide to do?

We want to point out that this is absolutely not the *only* such example, though! Try to come up with others!

Right now, all we have left to do is take the final diagram we constructed and use it to *define* an example and then prove it works. Here we go!

Proof. Define $A = \{1, 2, 3\}$ and $B = \{\star, \square\}$.

Define $f : A \to B$ by setting $f(1) = \star$, and $f(2) = \square$, and $f(3) = \star$.

Define $S = \{1, 2\}$ and $T = \{2, 3\}$.

Observe that $S \cap T = \{2\}$, so $\text{Im}_f(S \cap T) = \{f(2)\} = \{\square\}$.

However, observe that $\text{Im}_f(S) = \text{Im}_f(T) = B$, so $\text{Im}_f(S) \cap \text{Im}_f(T) \neq \{\square\}$.

Since $\star \in \text{Im}_f(S) \cap \text{Im}_f(T)$ but $\star \notin \text{Im}_f(S \cap T)$, this proves our claim. □

We have now seen an example of how to **prove** a claim about arbitrary functions and images, as well as how to **construct** a specific counterexample to **disprove** a claim. In the exercises, you will be asked to solve similar problems. Sometimes, you will need to *figure out* whether a claim is True or not. (Here, we *told* you which claim was valid beforehand.) We recommend trying one of two things: (1) Try to prove the claim, and see if it breaks down somewhere, or (2) Try to construct a counterexample, and see if you have trouble doing so. If you complete either task . . . well, hey, you figured it out! But if you're struggling, it might help you figure out what's really going on.

Specifically, you will be asked to examine the claim we discussed above, but with "∪" instead of "∩". What do you think will happen? Go ahead and try it!

7.3.3 Pre-Image: Definition and Examples

A natural question you might have now is: What about going the other way? Can we take a subset of the *codomain* and identify the elements whose outputs "land" in that set? Of course! This next definition provides us a term for this notion, and you'll notice many similarities with the definition of *image*.

Definition

Definition 7.3.8. *Let A, B be sets and let $f : A \to B$ be a function. Let $Y \subseteq B$.*

The **pre-image of Y under the function** *f is written and defined as*

$$\text{PreIm}_f(Y) = \{a \in A \mid f(a) \in Y\}$$

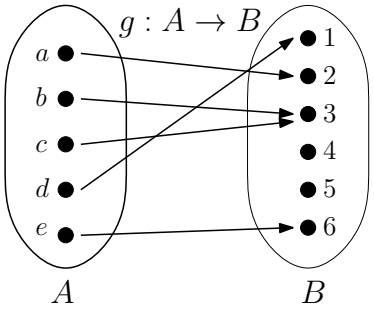
That is, the pre-image of Y under f is the set of all "inputs" that produce an "output" in Y.

(We will sometimes abbreviate the notation as just $\text{PreIm}(Y)$*, when the function is clearly identified and unambiguous, and consequently refer to the set as just "the pre-image of Y", instead of "the pre-image of Y under f".)*

Think about this first: What is $\text{PreIm}_f(B)$, where B is the entire codomain? Look back at the definition: this is the set of all inputs (in A) whose outputs "land" in B. That's all of A, of course, since f is a well-defined function! Accordingly, we will really only be working with sets $Y \subset B$, since those cases are more interesting.

Examples

Example 7.3.9. This first example uses the same function we defined in the last section when we discussed images. We'll show you the schematic diagram again, but spare you from re-defining all the details of the function. (See Example 7.3.2 for the details.)



Define $Z_1 = \{1, 2, 3\}$ and $Z_2 = \{2, 3, 4\}$ and $Z_3 = \{4, 5, 6\}$.

Let's identify the following pre-images and explain them.

(1) $\text{PreIm}_g(\{1\}) = d$

This is because $g(d) = 1$ and no *other* $x \in A$ satisfies $g(x) = 1$.

(Note: We need to use *set brackets* here. "$\text{PreIm}_g(1)$" would make no sense.)

(2) $\text{PreIm}_g(\{4\}) = \varnothing$

This is because *no* $x \in A$ satisfies $g(x) = 4$

(3) $\text{PreIm}_g(Z_1) = \{a, b, c, d\}$

This is because $g(a) = 2$, $g(b) = g(c) = 3$, and $g(d) = 1$, but no other $x \in A$ satisfies $g(x) \in Z_1$.

(4) $\text{PreIm}_g(Z_2) = \{a, b, c\}$

This is because $g(a) = 2$ and $g(b) = g(c) = 3$, but no other $x \in A$ satisfies $g(x) \in Z_2$.

(5) $\text{PreIm}_g(Z_3) = \{e\}$

This is because $g(e) = 6$, but no other $x \in A$ satisfies $g(x) \in Z_3$.

(6) $\text{PreIm}_g(\{5\}) = \varnothing$

This is because $\forall x \in A\text{.}\ g(x) \neq 5$.

Example 7.3.10. Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $\forall x \in \mathbb{R}\text{.}\ f(x) = x^2$.

Let's identify a few pre-images with this function. We will let *you* figure out why our claims are valid, as well as how to explain and prove them, this time!

(1) $\text{PreIm}_f(\{1\}) = \{-1, 1\}$

(2) $\text{PreIm}_f(\{y \in \mathbb{R} \mid y < 0\}) = \varnothing$

(3) $\text{PreIm}(\{y \in \mathbb{R} | y \geq 0\}) = \mathbb{R}$

(4) $\text{PreIm}(\{y \in \mathbb{R} | 0 < y < 1\}) = \{x \in \mathbb{R} \mid -1 < x < 1\}$

7.3.4 Proofs about Pre-Images

Notice that the following claim is one of **equality**. Compare this to Proposition 7.3.6, which has a similar statement about *images* but it is only a set *containment*. Interesting, right?

Proposition 7.3.11. *Let A, B be sets. Let $f : A \to B$ be a function. Let $X, Y \subseteq B$. Then,*

$$\text{PreIm}_f(X \cap Y) = \text{PreIm}_f(X) \cap \text{PreIm}_f(Y)$$

Notice how the proof below appeals directly to the formal *definition* of pre-images. We will jump right in and prove both parts. The exercises will ask you to investigate this claim with "$\cup$" instead of "$\cap$".

Proof. Let $x \in \text{PreIm}_f(X \cap Y)$ be arbitrary and fixed.

By the definition of pre-image, this means $f(x) \in X \cap Y$. Accordingly, $f(x) \in X$ and $f(x) \in Y$.

Since $f(x) \in X$, this means that $x \in \text{PreIm}_f(X)$ (by the definition of pre-image). Similarly, since $f(x) \in Y$, this means that $x \in \text{PreIm}_f(Y)$.

Thus, by the definition of intersection, we can deduce that $x \in \text{PreIm}(X) \cap \text{PreIm}(Y)$.

This shows $\text{PreIm}(X \cap Y) \subseteq \text{PreIm}(X) \cap \text{PreIm}(Y)$.

Next, let $y \in \text{PreIm}(X) \cap \text{PreIm}(Y)$ be arbitrary and fixed.

By the definition of pre-image, this means $y \in \text{PreIm}_f(X)$ and $y \in \text{PreIm}_f(Y)$.

Since $y \in \text{PreIm}_f(X)$, we can deduce that $f(y) \in X$, by the definition of pre-image. Similarly, since $y \in \text{PreIm}_f(Y)$, we can deduce that $f(y) \in Y$.

By the definition of interesection, this tells us $f(y) \in X \cap Y$. Then, by the definition of pre-image, this tells us $y \in \text{PreIm}(X \cap Y)$.

This shows $\text{PreIm}(X \cap Y) \supseteq \text{PreIm}(X) \cap \text{PreIm}(Y)$.

By a double-containment proof, we have proven the claim. □

You might read through this and think, "How does one come up with a proof like this?" Well, there isn't a whole lot of ingenuity behind a result like this. All we did was appeal directly to definitions. Everything fell into place from there.

If you find yourself *stuck* while working on a problem, or you're just unsure of where to start ... just write down the relevant definitions. Try to apply them to the statement you're trying to prove. See what happens!

A Proof with Pre-Images and Images

Let's work on one result that involves *both* of the concepts we have introduce in this section. We will prove one containment and ask you to *disprove* the other one in the exercises.

Proposition 7.3.12. *Let A, B be sets. Let $f : A \to B$ be a function. Let $Y \subseteq B$. Then,*

$$\mathrm{Im}_f\big(\mathrm{PreIm}_f(Y)\big) \subseteq Y$$

Proof. Let $b \in \mathrm{Im}_f\big(\mathrm{PreIm}_f(Y)\big)$ be arbitrary and fixed.

By the definition of image, this means $\exists a \in \mathrm{PreIm}_f(Y)\text{.}\ f(a) = b$. Let such an a be given.

Since $a \in \mathrm{PreIm}_f(Y)$, this means $f(a) \in Y$, by the definition of pre-image.

Since $b = f(a)$ and $f(a) \in Y$, this means $b \in Y$.

This proves the claim. □

7.3.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What are the differences between **image** and **pre-image**?

(2) Suppose $f : A \to B$ is a function. What is $\mathrm{PreIm}_f(B)$?

(3) Suppose $g : \mathbb{R} \to \mathbb{R}$ is a function. Why is the expression $\mathrm{Im}_g(0)$ not a proper statement? What do you think the writer of such an expression meant?

(4) Say $f : A \to B$ is a function and $Y \subseteq B$. What does it mean if $\mathrm{PreIm}_f(B) = \varnothing$? Is this possible?

(5) Say $f : A \to B$ is a function and $X \subseteq A$. What does it mean if $\mathrm{Im}_f(A) = \varnothing$? Is this possible?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Let $h : \mathbb{R} - \{-1\} \to \mathbb{R}$ be defined by $\forall x \in \mathbb{R} - \{-1\}\,.\; h(x) = \frac{x}{1+x}$.

Prove that $\text{Im}_h(\mathbb{R} - \{-1\}) = \mathbb{R} - \{1\}$.

Then, define $P = \{y \in \mathbb{R} \mid y > 1\}$ and $U = \{y \in \mathbb{R} \mid y > -1\}$.

Prove that $\text{PreIm}_h(P) = U$.

(2) Let $f : A \to B$ be a function. Let $S, T \subseteq A$. For each of the following claims, **prove** it must hold, or disprove it by finding a **counterexample**.

(a) $\text{Im}_f(S \cup T) \subseteq \text{Im}_f(S) \cup \text{Im}_f(T)$

(b) $\text{Im}_f(S \cup T) \supseteq \text{Im}_f(S) \cup \text{Im}_f(T)$

(3) Let $f : A \to B$ be a function. Let $Y, Z \subseteq B$. For each of the following claims, **prove** it must hold, or disprove it by finding a **counterexample**.

(a) $\text{PreIm}_f(Y \cup Z) \subseteq \text{PreIm}_f(Y) \cup \text{PreIm}_f(Z)$

(b) $\text{PreIm}_f(Y \cup Z) \supseteq \text{PreIm}_f(Y) \cup \text{PreIm}_f(Z)$

(4) Look back at Proposition 7.3.12. Consider the *reverse* containment:

$$\text{Im}_f\big(\text{PreIm}_f(Y)\big) \supseteq Y$$

Disprove the claim that this holds for any function $f : A \to B$ and any $Y \subseteq B$ by constructing a specific counterexample and proving that it works.

7.4 Properties of Functions

7.4.1 Surjective (Onto) Functions

You might be wondering something by now ... If we can identify the image of the domain under a given function, why bother with a codomain that's any "larger" than that set? Sure $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is a fine function, but changing the codomain to just the nonegative real numbers doesn't really affect anything. It might even make it better, because nothing in the codomain is "missed" by the function! If you're thinking this way, then you have anticipated our next definition, which encapsulates precisely this property of a function: when the codomain and the image of the domain are the same set.

Definition

Definition 7.4.1. *Let A, B be sets and let $f : A \to B$ be a function. We say f is a* **surjective** *function if and only if* $\text{Im}_f(A) = B$.

Equivalently, we just say "f is surjective" (adjectival form), or that "f is a surjection" (nounal form).

(The word "onto" is a fairly commonly used synonym for this term, so we will mention it here but won't use it again. This is just in case you've seen this word somewhere else.)

Referring back to the definition of image, we can state this property equivalently in the form of a quantified statement:

$$f \text{ is surjective} \iff \forall b \in B \bullet \exists a \in A \bullet f(a) = b$$

That is, f is surjective if and only if every output has at least one *corresponding input.*

Think for a minute about why the second form of this definition is really the same as the first one. The property that $\text{Im}_f(A) = B$ is a statement about *sets*. We already know that, by definition, $\text{Im}_f(A) \subseteq B$ (nothing in the image can fall "outside" of the codomain), so this further property means that $B \subseteq \text{Im}_f(A)$, as well. This is precisely what the second form of the definition says: every element of the codomain satisfies the defining property of being an element of the image.

Also, notice that nothing about the definition says the a we find to correspond to a b must be unique! All this property requires is that, for every $b \in B$, we can identify *at least one* $a \in A$ that satisfies $f(a) = b$. There might be more than one, there might be exactly one. It doesn't matter, as long as there aren't *none*.

What does the property of being a *surjection* mean in terms of a schematic diagram? Since every element of the codomain is "hit" by the function, this means that every dot on the right-hand side of the schematic has an incoming arrow. (Remember: this type of heuristic language is fine to keep in mind—we are using it to help describe these concepts, after all—but this does *not* constitute a proof. Any sentence of this sort that you use in a proof should be accompanied by a more rigorous statement, using mathematical language and/or logical symbols.) Why would we care about such a property? In general, it can be difficult to declare *exactly* what the image of a function is, and we might (at first) be able to only declare what the codomain is. Proving that, in fact, *all* of the codomain elements are outputs of the function can be additiona, helpful information!

Negating the Definition

Typically, we will define a function and then ask: is this a surjection or not? If we believe a function *is* a surjection, we should prove that by showing the

codomain and image are the same set. If we believe it is *not* a surjection, we should prove that by finding a *counterexample*. Let's look at the logical negation of the statement that defines a surjective function:

$$\neg(\forall b \in B\text{.}\ \exists a \in A\text{.}\ f(a) = b) \iff \exists b \in B\text{.}\ \forall a \in A\text{.}\ f(a) \neq b$$

That is, to prove a function f is *not* a surjection, we must find an element of the codomain that is *not* an element of the image. This involves some scratch work and intuition to identify such a b. From there, we must somehow show that no possible a satisfies $f(a) = b$. We might argue this directly by taking an arbitrary $a \in A$ and explaining why $f(a) \neq b$. Alternatively, we might argue this by contradiction: assuming that there is an $a \in A$ such that $f(a) = b$, we seek a contradiction. Either of these appraoches is reasonable, and they are logically equivalent.

Examples

Let's see these techniques in action with a few examples. For some of them, we might be able to use some graphical intuition or try a few test cases to figure out a *guess*, but ultimately we need to settle in and *prove* some logical statements to validate our claims.

Example 7.4.2. Consider $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by $p(a, b) = ab$. Is p surjective? Yes, it is! It looks like we can just allow a to be 1, so that the function outputs whatever b is. Let's make this observation more formal with a proof:

Proof. Let $n \in \mathbb{N}$ be arbitrary and fixed. Define $(a, b) = (1, n)$.

Notice that $(1, n) \in \mathbb{N} \times \mathbb{N}$ and $p(1, n) = 1 \cdot n = n$.

Since n was arbitrary, this shows p is surjective. □

Example 7.4.3. Let C be the set of all cars in the United States. Let S be the set of all strings of letters and digits that are of length at most 7 (i.e. these are the *potential* strings you might see on a car's license plate).

Let $f : C \to S$ be defined by inputting a car and outputting its license plate string. Is the function f a surjection?

No, definitely not! In case you weren't aware, *curse words* are disallowed on license plates! So certainly, there exist many strings of letters that you will *never* see on a license plate in the United States. (We'll let you provide some examples on your own . . .)

Because we have exhibited an element of S that is *not* an element of $\mathrm{Im}_f(C)$—or, at least, *you* thought of an example—we have shown that f is *not* a surjection.

Example 7.4.4. Let $d : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ be the function defined by

$$\forall (a, b) \in \mathbb{N} \times \mathbb{N}\text{.}\ d(a, b) = a - b$$

Let's determine whether d is a surjection and prove our claim. We might start by trying some "small values" for the input variables a and b. In the table below, the left column is a and the top row is b, and the entries are $d(a, b) = a - b$:

	1	2	3	4	5
1	0	-1	-2	-3	-4
2	1	0	-1	-2	-3
3	2	1	0	-1	-2
4	3	2	1	0	-1
5	4	3	2	1	0

It *looks* like all of the integers $z \in \mathbb{Z}$ will appear in this table. However, they don't all appear in one particular row or column. Rather, it looks like all the *non-negative* integers appear in the first column, while all the *non-positive* integers appear in the first row. Let's use these observations to write a proof. We'll take an arbitrary integer $z \in \mathbb{Z}$ and consider two **cases**; if $z \geq 0$, we will do one thing, and if $z < 0$, we will do something else. As long as we succeed in both cases, we will have proven that d is a surjection.

Proof. We claim d is a surjection. Let $z \in \mathbb{Z}$ be arbitrary and fixed. WWTS $\exists (a, b) \in \mathbb{N} \times \mathbb{N}.\ d(a, b) = z$. To do this, we consider two cases:

(1) Suppose $z \geq 0$. Then define $(a, b) = (z + 1, 1)$.

Since $z \geq 0$, we know $z + 1 \geq 1$ and so $z + 1 \in \mathbb{N}$. This guarantees $(z + 1, 1) \in \mathbb{N} \times \mathbb{N}$.

Also, notice that $d(z + 1, 1) = (z + 1) - 1 = z$.

(2) Suppose $z < 0$. Then define $(a, b) = (1, -z + 1)$.

Since $z < 0$, we know $-z > 0$ and so $-z + 1 \geq 2$, meaning $-z + 1 \in \mathbb{N}$. This guarantees $(1, -z + 1) \in \mathbb{N} \times \mathbb{N}$.

Also, notice that $d(1, -z + 1) = 1 - (-z + 1) = z$.

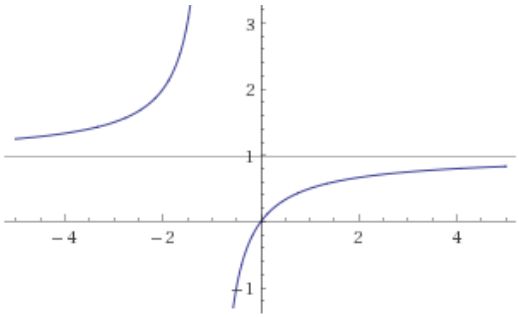
In either case, we are able to define $(a, b) \in \mathbb{N} \times \mathbb{N}.\ d(a, b) = z$. Since $z \in \mathbb{Z}$ was arbitrary, this proves that d is surjective. □

Example 7.4.5. Let $g : \mathbb{R} - \{-1\} \to \mathbb{R}$ be the function defined by

$$\forall x \in \mathbb{R}.\ g(x) = \frac{x}{1 + x}$$

(Notice why we have removed -1 from the domain. This ensures g is a *well-defined* function!)

Let's determine whether g is a surjection and prove our claim. As mentioned before, we can do some scratch work to figure out our claim: we could try plugging in some values of x, testing "extreme cases" by letting x get very close to -1 or letting x grow larger and larger ... All of this can help us plot a graph of the function, or we can just use some graphing software:



Regardless, none of this **proves** anything! What it does do is help us observe that this function g is **not** surjective. There seems to be a *horizontal asymptote* at $y = 1$. That is, the function g never "reaches" 1, but rather gets infinitely close. In terms of our new definition of *surjectivity*, this is decidedly a NO answer!

Try to prove this now. How can you show that the element $-1 \in \mathbb{R}$ is **not** an element of the image $\text{Im}_g(\mathbb{R})$? Try it! Then read on for our proof.

We will actually present *two* proofs here, for you to compare and contrast. They both accomplish the same goal—showing g is not surjective—but one does so by a **contradiction** method and the other by a **direct** method (using cases). Which do you think is better? Did you come up with one of these? Which is easier to read? We have no definitive opinion on these questions; they are both equally valid proofs!

Proof 1 (Direct). Let $x \in \mathbb{R} - \{-1\}$ be arbitrary and fixed. WWTS that $g(x) \neq 1$. We consider two cases:

- Suppose $x > -1$. This means $x + 1 > 0$, and so $\frac{1}{x+1} > 0$. We also know $x + 1 > x$ (which is true for every $x \in \mathbb{R}$.)

  By multiplying this inequality by the positive term $\frac{1}{x+1}$, we deduce that $1 > \frac{x}{x+1}$. Certainly, then, $g(x) = \frac{x}{x+1} \neq 1$.

- Suppose $x < -1$. This means $x + 1 < 0$, and so $\frac{1}{x+1} < 0$. We also know $x + 1 > x$.

  By multiplying this inequality by the negative term $\frac{1}{x+1}$ and switching the sign, we deduce that $1 < \frac{x}{x+1}$. Certainly, then, $g(x) = \frac{x}{x+1} \neq 1$.

In either case $g(x) \neq 1$. These cases cover all possibilities because $x \in \mathbb{R} - \{-1\}$ was arbitrary (and we need not consider $x = -1$). This shows

$$1 \notin \text{Im}_g(\mathbb{R} - \{-1\})$$

so g is not a surjection. □

Notice that this first proof proves an interesting *qualitative* observation about the graph: that the function lies above the horizonatal asymptote to the left of $x = -1$ and above the asymptote to the right of $x = -1$.

Proof 2 (Contradiction). AFSOC that g is surjective. This means

$$\forall y \in \mathbb{R}\text{.}\ y \in \text{Im}_g(\mathbb{R} - \{-1\})$$

In particular, then, we know $1 \in \text{Im}_g(\mathbb{R} - \{-1\})$, so $\exists x \in \mathbb{R} - \{-1\}\text{.}\ g(x) = 1$. Let such an x be given.

This means $g(x) = \frac{x}{x+1} = 1$. Multiplying both sides, we find $x = x + 1$. Subtracting, we find $0 = 1$, clearly a contradiction ⨳

Therefore, $1 \notin \text{Im}_g(\mathbb{R} - \{-1\})$, so g is not a surjection. □

Notice that this second proof does prove that g is not a surjection, but it doesn't add any other information about how the function behaves (like the previous proof did).

Let's move on from surjections and talk about a closley related property of functions.

7.4.2 Injective (1-to-1) Functions

When trying to prove a function is surjective, we took an arbitrary element of the codomain and had to find *at least one* element of the domain that corresponded to the original element. Sometimes there is *exactly one* such element, sometimes there are *many*, and sometimes there are *none*. What we will do now is consider those functions that fall into the "*exactly one*" case. We won't be presuming here that functions are already surjective. Rather, we are imposing this condition: we want there to be *no more than one* input for any given output. There might be exactly one or there might be none, but there certainly aren't two or more. These types of functions are special enough that we give them a name.

Definition

Definition 7.4.6. *Let A, B be sets and let $f : A \to B$ be a function. We say f is an* **injective** *function if and only if it has the property that*

$$\forall a_1, a_2 \in A\text{.}\ a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$$

Equivalently, we just say "f is injective" (adjectival form), or that "f is an injection" (nounal form).

(The term "1-to-1"—sometimes written "1-1"—is a fairly commonly used synonym for this word, so we will mention it here but won't use it again. This is just in case you've seen this term somewhere else.)

In other words, this defining property requires that "distinct inputs yield distinct

outputs". Also, remembering that the contrapositive *of a statement is logically equivalent, we can express this property as*

$$\forall a_1, a_2 \in A \,.\, f(a_1) = f(a_2) \implies a_1 = a_2$$

This expresses the equivalent notion that "if two outputs are equal, they must come from the same input".

Think about how this definition conveys the notion we described above. Say we have an injective function $f : A \to B$, and let's say we are given an element $b \in B$. Does this definition say that there is *at most one* element $x \in A$ such that $f(x) = b$? What possibilities does the definition allow?

Motivation

Let's motivate this by a particular application of functions. Think of a function as a *code-word machine* for you to send and receive secret messages with a friend. Your friend writes down a secret message, puts it in the encoder, and out pops a scrambled code that he sends to you. Later, you receive this scrambled code. You would *really* like to know that this code only came from *at most one* input phrase. What if you try to decode it and it comes out with both `I HATE YOU` **and** `I LOVE YOU`? What are you supposed to think then? Did your friend mean to sedn you both messages? What a terrible code system you've designed if both of those conflicting messages are encoded as the same scrambled message! In this context, it would be much nicer to have an encoding function where two *distinct* inputs can't possibly give the *same* output. This is presicely the defining property of being an injection.

Negating the Definition

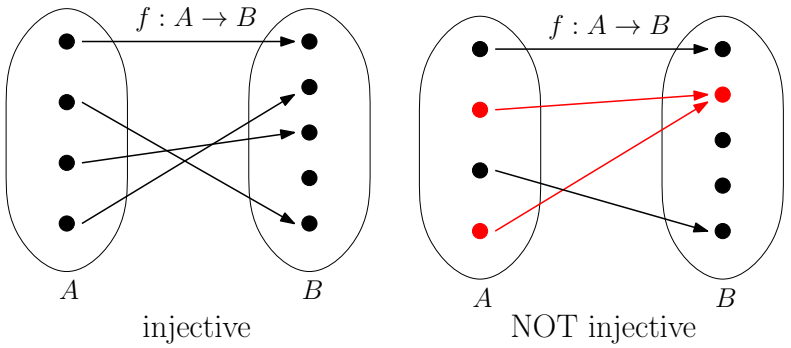
It might be helpful to think about the property of being an *injection* in terms of a schematic diagram, and in terms of the **negation** of the definition. Let's find that negation first:

$$\begin{aligned}\neg\big(\forall a_1, a_2 \in A \,.\, a_1 \neq a_2 \implies f(a_1) \neq f(a_2)\big) \\ \iff \big(\exists a_1, a_2 \in A \,.\, a_1 \neq a_2 \wedge f(a_1) = f(a_2)\big)\end{aligned}$$

(Remember that the negation of $P \implies Q$ is $P \wedge \neg Q$!)

This says a function is *not* injective if and only if we can find two *distinct* domain elements that output the *same* codomain element.

With that in mind, here are canonical examples of an injective and non-injective function:



The non-injective function has two distinct domain elements that output the same codomain element, whereas the injective function avoids this situation. It might feel a little odd to phrase a property in this kind of *negative* sense—a function is only injective if it *doesn't* have ...—but this is actually somewhat common in mathematics. (We will even see this idea later on when we talk about *infinite* sets, which are just ... sets that are *not* finite!) This negative formulation is easy enough to remember, and we can always relate it to another, positive formulation: an injective function has only 0 or 1 inputs corresponding to *any* given output.

Examples

Let's think about *how* to prove/disprove the injectivity of functions. As you might guess, the first two versions of the definition given above are useful when trying to show a function *is* injective: take two distinct elements of the domain and show their outputs are different, or take two equal outputs and show they came from equal inputs. The negation can also be used to show a function is injective via a proof by contradiction. Also, the third version is useful when proving a function is *not* injective: a counterexample amounts to finding two distinct inputs with the same output.

Let's see these techniques in action with a few examples. In fact, we will use some of the same examples we looked at in the previous section about surjections!

Example 7.4.7. Consider $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by $p(a, b) = ab$. Is p injective?

By trying some particular values of (a, b), we can see that p is definitely not an injection. Pick any number that has two different factorizations, like $12 = 3 \cdot 4 = 2 \cdot 6$. By letting $(a, b) = (3, 4)$ and $(c, d) = (2, 6)$, we can easily prove this claim. But we can do this even more easily, by noting that the *order* of the coordinates of an element like (a, b) *matters*!

Proof. This function is not injective. Let $(a, b) = (1, 2)$ and $(c, d) = (2, 1)$. Notice that $(a, b) \neq (c, d)$ because $1 \neq 2$. Also, notice that $p(a, b) = 1 \cdot 2 = 2$ and $p(c, d) = 2 \cdot 1 = 2$. Thus, $p(a, b) = p(c, d)$. This shows that p is not injective. □

Example 7.4.8. Let C be the set of all cars in the United States. Let S be the set of all strings of letters and digits that are of length at most 7 (i.e. these are the *potential* strings you might see on a car's license plate).

Let $f : C \to S$ be defined by inputting a car and outputting its license plate string. Is the function f an injection?

No, we don't think so! The same license plate string could appear on *different* cars that are registered in different *states*. Now, we don't have any examples of this on hand, so this isn't a totally formal proof, but hopefully you see the idea.

Could we amend the function definition to *make* it an injection? Sure, we could try! Consider also defining S to be the set of U.S. states. Let the function $g : C \to L \times S$ be defined by inputting a car and outputting the order pair of that car's license plate string and home state. This *will* be an injection, because no two cars in the same state can have the same plate. (Again, this is not really a formal proof; we are just trying to illustrate the concept of injectivity with a non-numerical example.)

Example 7.4.9. Let $d : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ be the function defined by $d(a, b) = a - b$. Determine whether d is an injection and prove your claim.

It turns out d is not an injection! Notice that $a - b = (a + 1) - (b + 1)$. We can use this to find a counterexample:

Consider the pairs $(2, 1) \in \mathbb{N} \times \mathbb{N}$ and $(3, 2) \in \mathbb{N} \times \mathbb{N}$. Notice that $d(2, 1) = 1$ and $d(3, 2) = 1$. Since $(2, 1) \neq (3, 2)$ and yet $d(2, 1) = d(3, 2)$, we conclude that d is not an injection.

Example 7.4.10. Let $F : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{Z})$ be defined by

$$\forall X \in \mathcal{P}(\mathbb{N}) \text{ . } F(X) = \bigcup_{a \in X} \{a, -a\}$$

Do you see what this function does? (Can you explain why it's even a *well-defined* function?)

Let's show you a few examples to give you an idea:

$$\begin{aligned} F\big(\{1\}\big) &= \bigcup_{a \in \{1\}} \{a, -a\} = \{-1, 1\} \\ F\big(\{1, 3, 5\}\big) &= \bigcup_{a \in \{1,3,5\}} \{a, -a\} = \{-1, 1\} \cup \{-3, 3\} \cup \{-5, 5\} \\ &= \{-5, -3, -1, 1, 3, 5\} \\ F(\varnothing) &= \bigcup_{a \in \varnothing} \{-a, a\} = \varnothing \\ F(\mathbb{N}) &= \mathbb{Z} - \{0\} \end{aligned}$$

We claim that F is an injection. Think about how to prove this before reading our proof. In particular, think about the different strategies we might employ here, based on the formal *definition* of injectivity. Might one strategy be more fruitful than another?

Proof. WWTS F is an injection. Let $X, Y \in \mathcal{P}(\mathbb{N})$.

Suppose that $X \neq Y$. WWTS $F(X) \neq F(Y)$.

Since $X \neq Y$, we have two cases: either $X \not\subseteq Y$ or $Y \not\subseteq X$ (or both).

Suppose $X \not\subseteq Y$. This means $\exists n \in X \,.\, n \notin Y$. Let such an n be given.

Since $n \in \{-n, n\}$ and $n \in X$, we see that $n \in F(X)$, by the definition of F.

However, since $n \notin Y$, we see that $\forall a \in Y \,.\, n \notin \{-a, a\}$. This follows because $n \notin Y$, as well as the fact that $n \in \mathbb{N}$ and $Y \subseteq \mathbb{N}$, so $\forall a \in Y \,.\, n \neq -a \in \mathbb{Z}$.

Accordingly, $n \notin F(Y)$. This shows that $F(X) \neq F(Y)$.

In the other case, where $Y \not\subseteq X$, we can follow the exact same argument with the roles reversed (i.e. switching X and Y in every step). This shows that $F(Y) \neq F(X)$.

Together, we have shown that $\forall X, Y \in \mathcal{P}(\mathbb{N}) \,.\, X \neq Y \implies F(X) \neq F(Y)$. This shows F is an injection. □

Think about how this proof might go if we used a different technique. Say we started by assuming $X, Y \in \mathcal{P}(\mathbb{N})$ and that $F(X) = F(Y)$. Can we deduce that $X = Y$?

7.4.3 Proof Techniques for Jections

Let's summarize the concepts of this section so far by presenting some **proof templates**. These can be used when you are trying to prove/disprove that a function is injective/surjective. We like using the shorthand "Jections" to refer to these two function properties together.

Prove that f is surjective

1. Let $b \in B$ be arbitrary and fixed.
2. Define $a =$ _____.
3. Show that $a \in A$.
4. Show that $f(a) = b$.
5. This shows that $b \in \text{Im}_f(A)$. Thus, $\text{Im}_f(A) = B$, so f is surjective.

Prove that f is not surjective

1. Define $b =$ _____.
2. Show that $b \in B$.
3. Let $a \in A$ be arbitrary and fixed.

4. Show that $f(a) \neq b$.

   (Alternatively, suppose $f(a) = b$ and find a contradiction.)

5. This shows that $\exists b \in B \centerdot\ b \notin \mathrm{Im}_f(A)$, so f is not surjective.

Prove that f is injective

1. Let $x, y \in A$ be arbitrary and fixed.
2. Suppose that $f(x) = f(y)$.
3. Deduce that $x = y$.

Alternatively:

1. Let $x, y \in A$ be arbitrary and fixed.
2. Suppose that $x \neq y$.
3. Deduce that $f(x) \neq f(y)$.

Prove that f is not injective

1. Define $x =$ _____ and define $y =$ _____.
2. Show that $x \in A$ and $y \in A$.
3. Show that $x \neq y$.
4. Show that $f(x) = f(y)$.

Prove that f is bijective

1. Prove that f is injective.
2. Prove that f is surjective.

7.4.4 Bijections

You might have guessed what we have been building towards here. Think about the two main properties of functions we just studied: *surjectivity* and *injectivity*. What happens when a function has *both* of these properties? What if a function has the property that, for every element of the codomain, there is *at least one* corresponding element in the domain (surjectivity) *and* there is also *at most one* such element (injectivity)? That's right: for every output, there is *exactly* one input! This is an incredibly nice property to have, and will be the foundation for our forthcoming discussion of *cardinality* (i.e. the *size* of a set). Let's make a definition and then discuss some examples.

Definition

Definition 7.4.11. *Let A, B be sets and let $f : A \to B$ be a function. We say f is a* **bijective** *function if and only if f is* both *injective and surjective.*

Equivalently, we just say "f is bijective" (adjectival form), or that "f is a bijection" (nounal form).

We will sometimes say that f is a bijection between *the sets A and B, instead of saying "from A to B". (The reason for this will become clear in the next section!)*

Notice that this definition is, logically speaking, an AND statement. For the moment, anyway, the only technique we have to *prove* a function is bijective is to just prove it is surjective *and* prove it is injective. Similarly, to prove a function is not bijective, we need to prove it is *either* not surjective or not injective. (It might be that both properties fail, but one such proof is sufficient to show a function is not bijective.) Rather than go over these same techniques (which are nicely summarized right before this section), we will just point out whether some of the examples we have seen thus far are bijections are not.

Example 7.4.12.

(a) Let $p : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be the function defined by $p(a, b) = ab$.

We proved that p is surjective but *not* injective, so it is **not** a bijection.

(b) Let $d : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ be the function defined by $d(a, b) = a - b$.

We proved that d is surjective but *not* injective, so it is **not** a bijection.

(c) Let $g : \mathbb{R} - \{-1\} \to \mathbb{R}$ be the function defined by

$$\forall x \in \mathbb{R} - \{-1\}.\ g(x) = \frac{x}{1+x}$$

We proved that g is not surjective. (Specifically, we showed $1 \notin \text{Im}_g(\mathbb{R} - \{-1\})$.) We will ask you in this section's exercises to prove that g *is* an injection, though. Together this means g is not a bijection.

However, consider defining $h : \mathbb{R} - \{-1\} \to \mathbb{R} - \{1\}$ by the same "rule" as g, i.e. $\forall x \in \mathbb{R} - \{1\}.\ h(x) = \frac{x}{1+x}$.

We asked you to prove in the exercises of Section 7.3.5 that this function satisfies $\text{Im}_h(\mathbb{R} - \{-1\}) = \mathbb{R} - \{1\}$. This shows that h is a surjection.

Furthermore, we will ask you to prove in this section's exercises that a function defined in this way—by taking an injection, using the same "rule", and redefining the codomain to be the image—produces a bijection.

Together, all of this proves that h is a **bijection** from $\mathbb{R} - \{-1\}$ to $\mathbb{R} - \{1\}$.

Example 7.4.13. Let's look at one new example, specifically chosen to preview some of the main ideas coming up ahead. Define $E \subseteq \mathbb{N}$ to be the set of all *even*

natural numbers; that is,

$$E = \{e \in \mathbb{N} \mid \exists k \in \mathbb{N}.\ e = 2k\}$$

Define the function $d : \mathbb{N} \to E$ by $d(n) = 2n$. We claim d is a bijection.

Proof. First, let's prove d is a surjection. Let $e \in E$ be given.

By the definition of E, $\exists k \in \mathbb{N}$ such that $e = 2k$. Let such a k be given.

This tells us $d(k) = 2k = e$. Since e was arbitrary, we conclude that d is a surjection.

Second, let's prove d is an injection. Let $m, n \in \mathbb{N}$ and assume $d(m) = d(n)$.

This means $2m = 2n$. Canceling the 2s from both sides, we find that $m = n$. Thus, d is an injection.

Together, this proves that d is a bijection. □

We'll motivate some future considerations by posing some questions: Does it seem a little strange to you that there is a *bijection* between $\mathbb{N}$ and E, a set that is a *proper* subset of $\mathbb{N}$? Is it always possible to find a bijection between a set and a subset of itself? Have we seen other examples of this situation before?

Motivation

The main idea behind a bijection $f : A \to B$ is that we can **pair up** the elements of A and B and identify them with each other, one by one. This idea follows from the definitions of both surjectivity and injectivity: every output has *exactly one* corresponding input. Furthermore, think more carefully about what we show in the *proofs* of such properties. In proving f is surjective, we show we can "move" from the codomain back to the domain in at least one way, and then in proving f is injective, we show that this is the *only* way to do it. In a sense, we are showing how to "undo" the function f and reverse its action. In fact, we are implicitly defining a new function from B back to A. Have you previously talked about the *inverse* of a function? That is precisely what we are rediscovering now! To make this notion of "moving back from the codomain to the domain" rigorous enough, we need to have a brief discussion about how to "combine" functions appropriately. Right after that, we will be able to give a precise definition of what we mean by the *inverse* of a function, and relate this to bijections. All of this happens in the next section.

7.4.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you

can confidently answer these before moving on will help your understanding and memory!

(1) Write down a definition of **surjective** in terms of an **image**. Then, write down a definition of surjective in terms of quantifiers.

(2) Describe two different ways of proving that a function is **injective**.

(3) Can a function be both injective and surjective? If so, give an example.

(4) Can a function be neither injective nor surjective? If so, give an example.

(5) Consider the following schematic diagrams. For each one, declare whether or not it is a function; and, if it is, declare whether or not it is (a) an injection and (b) a surjection.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Suppose $f : \mathbb{R} \to \mathbb{R}$ is an *increasing* function; that is, suppose

$$\forall x, y \in \mathbb{R}\text{.}\ x < y \implies f(x) < f(y)$$

Prove that f must be *injective*.

Then, prove that f *need not* be surjective by defining an increasing function that is not surjective.

(2) Let $g : \mathbb{R} - \{-1\} \to \mathbb{R}$ be the function defined by

$$\forall x \in \mathbb{R}\text{.}\ g(x) = \frac{x}{1+x}$$

Is g injective or not? *Prove* your claim.

(3) Give an example of a function $f : \mathcal{P}(\mathbb{N}) \to \mathbb{N}$ that is surjective. *Prove* that it is.

(**Hint:** Be careful about the fact that $\varnothing \in \mathcal{P}(\mathbb{N})$. Also, consider looking at Section 5.5.2 for some inspiration . . .)

(4) Give an example of a function $F : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ that is injective. *Prove* that it is.

Then, *prove* that your function F is *not* surjective.

(Note: Yes, we are asking you to prove your function is not surjective *without knowing what function you defined*. We know we are right! You will learn about our trick later in this chapter . . .)

(5) Suppose $f : A \to B$ and $g : B \to C$ are surjective functions. Prove that $g \circ f : A \to C$ is also surjective.

(6) Let $f : A \to B$ be an injective function. Define $g : A \to \mathrm{Im}_f(A)$ by setting $\forall x \in A \,.\, g(x) = f(x)$. Prove that g is a bijection.

(7) Define $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ by $F(x, y) = (x + y, 2x - y)$. Prove F is bijective.

(**Hint:** In your scratch work, you should try to solve a system of two equations. See Section 1.3.2 for some suggestions about how to do that.)

(8) Let A, B be sets. Let $g : A \to B$ be an *injection*.

Let $X \subseteq A$. Let $h : X \to B$ be the function defined by $\forall x \in X \,.\, h(x) = g(x)$. (That is, h is defined by the same "rule" as g, but on a "restricted domain".)

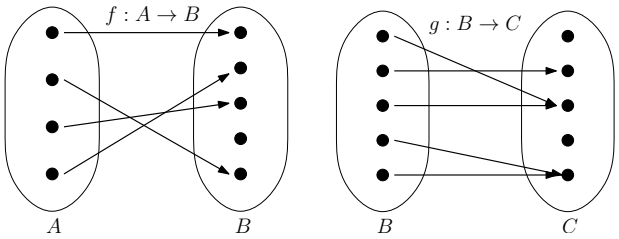
Prove that h is also an injection.

7.5 Compositions and Inverses

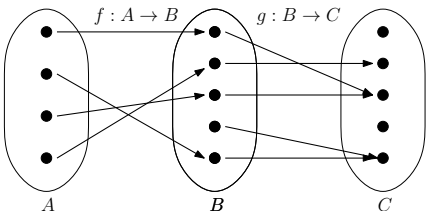
7.5.1 Composition of Functions

Motivation

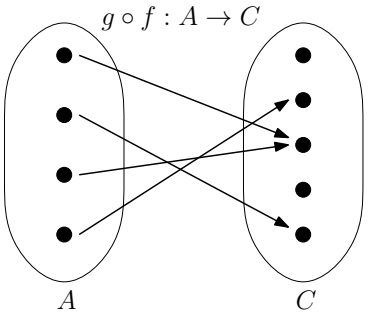
Let's think about the schematic interpretation of functions for a moment. Imagine that we have a function $f : A \to B$ and we also have a function $g : B \to C$, defined like this:



In a heuristic sense, f is like a "map" that gives us a particular route from elements of A to elements of B, while g is like a "map" from elements of B to elements of C. What would happen if we were to simply follow the "maps" one after the other? That is, let's combine the two by overlaying them,



and then simply travel from A all the way to C, cutting out the middle man:



This seems like a reasonable thing to do, right? Yes, of course it is! Whenever we have mathematical objects at our disposal, we're always curious about how we can resonably combine them and manipulate them and generalize them. In the case of functions, we call this combination a **composition** of functions. You might notice that such a composition really only makes sense if the codomain of the "first" function and the domain of the "second" function are the same. This is incorporated in the following definition.

Definition

Definition 7.5.1. *Let A, B, C be sets, and let $f : A \to B$ and $g : B \to C$ be functions. Consider the function $h : A \to C$ defined by*

$$\forall a \in A \bullet\ h(a) = g(f(a))$$

We say that h is the **composition** *of g with f and we write $h = g \circ f$.*

We also shorten this terminology and say h is "g composed with f".

This incorporates all the ideas we mentioned above. It requires that the codomain of f (the "first" function applied) to be the domain of g (the "second" function applied).

Another intuitive idea is to think of a function as a *machine* or a *black box*. Elements of the domain go in and elements of the codomain come out. We don't necessarily know what the machine does; we only see what comes out. Now,

think of hooking up two machines, one for f and one for g; take the output of f's machine and plug it into g's machine. What comes out is an element of C. We can take the work of these two machines and think of it as the work of one bigger machine. This is what the *composition* $g \circ f$ does; it's one larger machine that takes the operations of two machines and does them in a specified order.

Notation

Notice the ordering of the notation $g \circ f$ and how it compares to the order in which we *apply* the functions: f comes first, and then g, i.e. $g(f(a))$. In words, we would read "$g(f(a))$" as "g of f of a". In fact, if you find yourself having trouble remembering this order, here's a recommendation: read the "$\circ$" out loud as "after". Thus, $h = g \circ f$ would mean "g *after* f", because we take an element of a, apply f first, and *then* apply g.

It is also important to remember the *notation* of composed functions and to distinguish the function $g \circ f$ itself from an *application* of the function $g \circ f$ to some element $x \in A$. For instance, to write "g of f of x" using the "$\circ$" notation, we would write

$$(g \circ f)(x)$$

because we are "hitting" the element x with the function $g \circ f$. However, the following notation **make no sense** because it mixes up the ideas of functions and elements:

$$g \circ f(x)$$

Do you see the difference? The object $f(x)$ is an element of B, the codomain of f. But g is a function. What does it mean to compose a function with an element of a set? This doesn't work. Be careful with this, in general! This distinction will be especially important when we have to compose several functions together, like $(h \circ (g \circ k) \circ f)(z)$, where z is an element of f's domain, and f, g, h, k are functions.

Examples

Example 7.5.2. Let $C : \mathbb{R} \to \mathbb{R}$ be defined by

$$\forall x \in \mathbb{R}\text{.}\ C(x) = x - 273.15$$

Let $F : \mathbb{R} \to \mathbb{R}$ be defined by

$$\forall x \in \mathbb{R}\text{.}\ F(x) = \frac{9}{5}x + 32$$

The function C converts a temperature from degrees Kelvin to degrees Celsisus. The function F converts from degrees Celsius to degrees Fahrenheit.

Then the function $F \circ C$ converts from degrees Kelvin to degrees Fahrenheit

directly. We can compose the "rules" for the functions and find a formula for this direct conversion:

$$\begin{aligned}\forall x \in \mathbb{R}.\ (F \circ C)(x) &= F(C(x)) = F(x - 273.15) \\ &= \frac{9}{5} \cdot \big(x - 273.15\big) + 32 = \frac{9}{5}x - 459.67\end{aligned}$$

Example 7.5.3. Let $f : \mathbb{R} \to \mathbb{Z}$ be the function defined by

$$\forall x \in \mathbb{R}.\ f(x) = \lfloor x \rfloor$$

(Recall that $\lfloor x \rfloor$ is the *floor* of x: it is the *largest* integer $z \in \mathbb{Z}$ that satisfies $z \leq x$. Let $g : \mathbb{Z} \to \mathbb{N}$ be the function defined by

$$\forall z \in \mathbb{Z}.\ g(z) = \begin{cases} -z & \text{if } z < 0 \\ z + 1 & \text{if } z \geq 0 \end{cases}$$

Let's find $g \circ f$. Notice that whenever $x \in \mathbb{R}$ satisfies $x < 0$, we will have $\lfloor x \rfloor x < 0$, as well. Similarly, whenever $x \in \mathbb{R}$ satisfies $x \geq 0$, we will have $\lfloor x \rfloor \geq 0$. This tells us that the composition $g \circ f$ will also be a **piece-wise** function:

$$\forall x \in \mathbb{R}.\ (g \circ f)(x) = \begin{cases} -\lfloor x \rfloor & \text{if } x < 0 \\ \lfloor x \rfloor + 1 & \text{if } x \geq 0 \end{cases}$$

Questions: Is this function injective? Surjective? Try to prove your claims!

Example 7.5.4. Define $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$ and $h : \mathbb{N} \to \mathbb{N}$ by

$$\begin{aligned}&\forall n \in \mathbb{N}.\ f(n) = n + 3 \\ &\forall n \in \mathbb{N}.\ g(n) = n^2 \\ &\forall n \in \mathbb{N}.\ h(n) = 2n - 1\end{aligned}$$

(Question: Are you sure these are well-defined functions? Why?)

We can find "rules" for the compositions $g \circ f$ and $h \circ f$:

$$\begin{aligned}&\forall n \in \mathbb{N}.\ (g \circ f)(n) = g(f(n)) = g(n + 3) = (n + 3)^2 = n^2 + 6n + 9 \\ &\forall n \in \mathbb{N}.\ (h \circ g)(n) = h(g(n)) = h(n^2) = 2n^2 - 1\end{aligned}$$

We can then use these to find a rule for a further composition, like $h \circ (g \circ f)$:

$$\begin{aligned}\forall n \in \mathbb{N}.\ \big(h \circ (g \circ f)\big)(n) &= h\big((g \circ f)(n)\big) = h\big(n^2 + 6n + 9\big) \\ &= 2(n^2 + 6n + 9) - 1 = 2n^2 + 12n + 17\end{aligned}$$

Likewise, we can use these to find a rule for $(h \circ g) \circ f$:

$$\begin{aligned}\forall n \in \mathbb{N}.\ \big((h \circ g) \circ f\big)(n) &= \big(h \circ g\big)(f(n)) = (h \circ g)(n + 3) \\ &= 2(n + 3)^2 - 1 = 2(n^2 + 6n + 9) - 1 \\ &= 2n^2 + 12n + 17\end{aligned}$$

Look at that, they're the same rule! That is, we just *proved* that

$$(h \circ g) \circ f = h \circ (g \circ f)$$

in the sense of *functions* by showing that they yield the same output on *every* allowable input.

Composition is Associative

There was nothing particularly special about the functions f, g, h used in the previous example. The result we obtained is actually true *in general*. The following theorem and its proof will show this. We are proving that function composition is **associative**. This means that whenever we have a string of compositions, we can move the parentheses around at will; we know that the order in which we apply the parentheses doesn't matter.

Theorem 7.5.5. *Let A, B, C, D be any sets. Let $f : A \to B$ and $g : B \to C$ and $h : C \to D$ be functions. Then,*

$$h \circ (g \circ f) = (h \circ g) \circ f$$

Proof. WWTS that the outputs of the two functions $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are the same, for every possible input.

Let $x \in A$ be given. Applying the definition of *composition*, we see that

$$[h \circ (g \circ f)](x) = h\big(g \circ f)(x)\big) = h\big(g(f(x))\big)$$

and

$$[(h \circ g) \circ f](x) = \big(h \circ g\big)(f(x)) = h\big(g(f(x))\big)$$

□

Compositions and Jections

Here's something interesting to ponder now: What happens if we take the composition of two functions with a shared property? Does that property "carry over", as well? For instance, if we compose two injections, do we get another injection? Does only *one* of the composed functions *need* to be an injection to guarantee the composition is an injection?

Similarly, let's say we have a composition of two functions. If we know the composition is a surjection, can we *necessarily* deduce that one of the functions we composed is also a surjection? Do they both need to be?

We will state and prove some claims about questions like these in this short section. We will let you prove some related facts (or find appropriate counterexamples, as the case may be) in the exericses, both for this section and at the end of the chapter.

Proposition 7.5.6. *Let A, B, C be sets and let $f : A \to B$ and $g : B \to C$ be functions. If $g \circ f$ is injective, then f is necessarily injective.*

(Notice that this doesn't assume *any* properties of g; it doesn't even have to be injective, necessarily! As an exercise, try to find an example of functions $f : A \to B$ and $g : B \to C$ such that $g \circ f$ is injective and g is injective, and also an example where $g \circ f$ is injective but g is not injective.)

Proof. Let $x, y \in A$ be given. Suppose $f(x) = f(y)$. WWTS $x = y$.

Since g is a well-defined function, $g(f(x)) = g(f(y))$.

This means $(g \circ f)(x) = (g \circ f)(y)$.

Since $g \circ f$ is injective, $x = y$. This was our goal, so the claim is proven. □

It turns out that the *converse* of the claim we just proved is False. Since that claim is one about *all* functions, disproving it requires us to produce a counterexample.

Proposition 7.5.7. *Let A, B, C be sets and let $f : A \to B$ and $g : B \to C$ be functions. Suppose f is injective. Then it is not necessarily the case that $g \circ f$ is injective.*

Try doing some scratch work on your own to come up with a counterexample before reading about ours. Remember that you don't need to find the most interesting or complicated one, nor do you necessarily need one defined by a *rule*; you just need to be able to define one!

Proof. We will exhibit a counterexample.

Define $A = \{1, 2\}$ and $B = \{\heartsuit, \diamondsuit\}$ and $C = \{\star\}$.

Define f by setting $f(1) = \heartsuit$ and $f(2) = \diamondsuit$.

Notice f is injective because $f(1) \neq f(2)$.

Define g by setting $g(1) = g(2) = \star$.

Notice $g \circ f$ is defined by $(g \circ f)(1) = \star$ and $(g \circ f)(2) = \star$.

This shows $g \circ f$ is not injective, because $(g \circ f)(1) = (g \circ f)(2)$ but $1 \neq 2$. □

7.5.2 Inverses

Motivation

As we said before, a **bijection** $f : A \to B$ has a very nice property, in that f "pairs off" the elements of the two sets, A and B. Given an element $a \in A$, there is *exactly one* element $b \in B$ that satisfies $f(a) = b$. This is because f is a well-defined function. But we also know that a is the *only* domain element associated with b in this way. This is because f is a bijection. Because of this unique association in both directions, we can think of "reversing" the action of f. Given an element $b \in B$, identify the a that would produce that b. This is what an **inverse** function does. Here, we will define it in terms of function **composition** and **identity** functions. This is also the reason we say a bijection

is *between* two sets as opposed to just from one set to the other; as soon as we have it one way, we know we can have it the other way, too!

Before we see the definition, let's quickly recall the definition of the **identity** function that we saw before. It plays an important role in the forthcoming definition of inverse.

> **Definition:** Given a set X, the **identity function** $\mathrm{Id}_X : X \to X$ is defined by $\forall z \in X \,.\, \mathrm{Id}_X(z) = z$.

Definition

Notice that this definition doesn't say *anything* about the functions being bijections. This is purely a formal definition of what an inverse function means. Afterwards, we will have to prove any claims about hwo inverses and bijections are related.

Definition 7.5.8. *Let $f : A \to B$ be a function. Suppose there is a function $g : B \to A$ such that $f \circ g : A \to A$ satisfies $f \circ g = \mathrm{Id}_A$ and $g \circ f : B \to B$ satisfies $g \circ f = \mathrm{Id}_B$.*

Then we say g is the **inverse** *of f and write $g = f^{-1}$.*

(Notice that some conditions are implicitly stated by the assumptions and conclusions in the definition above. Specifically, it must be that $B = \mathrm{Im}_f(A)$, to make sure g is a function. Likewise, $A = \mathrm{Im}_g(B)$.)

Example

Let's look back at a function we saw before when we discussed bijections. With your help in the exercises, we learned that this function is a bijection. Here, we will find its inverse.

Example 7.5.9. Let $h : \mathbb{R} - \{-1\} \to \mathbb{R} - \{1\}$ be defined by

$$\forall x \in \mathbb{R} - \{-1\} \,.\, h(x) = \frac{x}{1+x}$$

To *find* a candidate function that will be the inverse of h, it usually helps to set the "rule" for h equal to some new variable, and then solve for x.

Here, let's say $h(x) = y$. How can we "reverse" this process and identify what x is, in terms of y? Observe that we can make some algebraic steps, as follows:

$$\begin{aligned} h(x) = y &\iff \frac{x}{1+x} = y \\ &\iff (1+x)y = x \\ &\iff xy + y = x \\ &\iff y = x(1-y) \\ &\iff x = \frac{y}{1-y} \end{aligned}$$

This *scratch work* has given us a candidate for the inverse of h. We haven't *proven* anything with these observations! What we have to do now is make a claim and then demonstrate, for the reader, all of the essential facts. Notice that we took care to define a *new* function H, and used it to **prove** that $H = h^{-1}$, in fact. It would be presumptuous and erroneous to **define** h^{-1} and then work with it. We are trying to show h has an inverse, so we can't just declare it has one at the beginning of our proof!

Proof. Define $S = \mathbb{R} - \{-1\}$ and $T = \mathbb{R} - \{1\}$ for convenient shorthand, so $h : S \to T$.

Let $H : T \to S$ be the function defined by $\forall y \in T \bullet H(y) = \frac{y}{1-y}$.

First, let's show that H is a well-defined function. For every $y \in T$, we know $y \neq 1$, so $1 - y \neq 0$. Thus, the fraction $\frac{y}{1-y}$ is a well-defined real number.

Furthermore, we can argue that $\frac{y}{1-y} \neq -1$. AFSOC that $\frac{y}{1-y} = -1$. Then multiplying through by $1 - y$ tells us $y = y - 1$, a clear contradiction.

Second, let's show that $H \circ h = \mathrm{Id}_S$. Let $x \in S$ be given. Observe that

$$\begin{aligned}(H \circ h)(x) &= H(h(x)) = H\left(\frac{x}{1+x}\right) \\ &= \frac{\frac{x}{1+x}}{1 - \frac{x}{1+x}} \cdot \frac{1+x}{1+x} = \frac{x}{(1+x)-x} \\ &= \frac{x}{1} = x\end{aligned}$$

Third, let's show that $h \circ H = \mathrm{Id}_T$. Let $y \in T$ be given. Observe that

$$\begin{aligned}(h \circ H)(y) &= h(H(y)) = h\left(\frac{y}{1-y}\right) \\ &= \frac{\frac{y}{1-y}}{1 + \frac{y}{1-y}} \cdot \frac{1-y}{1-y} = \frac{y}{(1+y)-y} \\ &= \frac{y}{1} = y\end{aligned}$$

Therefore, by the definition of inverse, $H = h^{-1}$. □

Checking Both Directions

Let's say $f : A \to B$ is a function, and you have made a claim about f having an inverse by defining a function $g : B \to A$. It is **extremely** important that you show **both** compositions yield the identity function; that is, you must show both

$$f \circ g = \mathrm{Id}_B \quad \text{and} \quad g \circ f = \mathrm{Id}_A$$

You might occasionally forget to do so, or you just might not see why this is necessary. To help you understand this importance, we have included Exercise

2 in Section 7.5.4 below. It asks you to find an example where "one way" yields the identity function but the "other way" does *not*, so that the proposed function is actually not an *inverse*. Try to find several examples, if you can. The more striking you make this point, the better!

7.5.3 Bijective $\iff$ Invertible

As we have been hinting at all along, a bijective function has an inverse. This claim's converse holds, as well, so we can state and prove this *if and only if* statement. The word in the section heading here—**invertible**—is often used to mean "has an inverse".

Theorem 7.5.10. *Let A, B be any sets. Let $f : A \to B$ be a function. Then,*

$$f \text{ is bijective} \iff f \text{ has an inverse } f^{-1} : B \to A$$

Proof. ($\implies$) Assume f is bijective. This means f is surjective and injective.

We need to define an inverse function for f. Let's define $g : B \to A$ as follows:

Let $b \in B$ be given. Since f is surjective, we know $\exists a \in A.\ f(a) = b$. Let such an a be given. Since f is injective, we know that

$$\forall x \in A.\ x \neq a \implies f(x) \neq f(a) = b$$

That is, we know this a is the *unique* element of A that satisfies $f(a) = b$. Let's define $g(b) = a$. This is a well-defined function because of these observations.

Next, observe that $(f \circ g)(b) = f(g(b)) = f(a) = b$, so $f \circ g = \mathrm{Id}_B$.

Also, observe that $(g \circ f)(a) = g(f(a)) = g(b) = a$, so $g \circ f = \mathrm{Id}_A$.

Therefore, $g = f^{-1}$, so f has an inverse.

($\impliedby$) Assume f has an inverse function, $f^{-1} : B \to A$.

First, let's show f is injective. Let $a_1, a_2 \in A$ be given. Observe that

$$\begin{aligned}
f(a_1) = f(a_2) &\implies f^{-1}(f(a_1)) = f^{-1}(f(a_2)) && f^{-1} : B \to A \text{ is a } \textit{function} \\
&\implies (f^{-1} \circ f)(a_1) = (f^{-1} \circ f)(a_2) && \text{definition of composition} \\
&\implies \mathrm{Id}_A(a_1) = \mathrm{Id}_A(a_2) && \text{definition of identity} \\
&\implies a_1 = a_2 && \text{definition of identity}
\end{aligned}$$

Thus, f is injective.

Second, let's show f surjective. Let $b \in B$ be given. Since f^{-1} is a function, we know $\exists a \in A.\ f^{-1}(b) = a$. Let such an a be given. Then observe that$f^{-1}(b) =$

a.

$$\begin{aligned} f^{-1}(b) = a &\implies f(f^{-1}(b)) = f(a) && f : A \to B \text{ is a function} \\ &\implies (f \circ f^{-1})(b) = f(a) && \text{definition of composition} \\ &\implies \mathrm{Id}_B(b) = f(a) && \text{definition of identity} \\ &\implies b = f(a) && \text{definition of identity} \end{aligned}$$

□

Proving a Function is Bijective

This helpful theorem now provides us with another technique for proving that a given function $f : A \to B$ is a bijection. Rather than proving f is an injection *and* a surjection, we can just define a new function $g : B \to A$ and prove that it is the **inverse** of f, i.e. $g = f^{-1}$. Then, this theorem applies and tells us that f is a bijection! Depending on the context, one or the other of these strategies might be easier to apply, or you might just be more comfortable with one of them. Keep in mind that both strategies are viable, though!

Inverse of an Inverse

The following corollary follows immediately from the theorem above. We call it a *corollary* and not its own theorem because it doesn't really assert anything amazingly new; rather, its conclusion comes from applying the theorem above, as you'll see in the proof.

Corollary 7.5.11. *Let A, B be any sets. Let $f : A \to B$ be a function.*

If f is a bijection, then f^{-1} exists and it is also *a bijection.*

Furthermore, $\left(f^{-1}\right)^{-1} = f$.

Proof. Suppose f has an inverse, $f^{-1} : B \to A$. This means $f \circ f^{-1} = \mathrm{Id}_B$ and $f^{-1} \circ f = \mathrm{Id}_A$, by the definition of inverse.

These are precisely the conditions that show $\left(f^{-1}\right) = f$, again by the definition of inverse! This shows f^{-1} has an inverse (namely, f itself) so the theorem above tells us that f^{-1} must be a bijection. □

Inverse of a Composition

Before we move on to some exercises and the next section, let's get your help in putting together the main ideas of this chapter so far. Specifically, we are going to state two results here. The proofs are left for you in the chapter exercises. By working through those proofs, you will (a) solidify your understanding of many of the concepts introduced so far—functions, jections, compositions, inverses—and (b) obtain a helpful result about how to define the inverse of a composition of functions!

Proposition 7.5.12. *Let $f : A \to B$ and $g : B \to C$ be bijective functions. Define $h : A \to C$ to be $h = g \circ f$. Then h is also bijective.*

Proof. Left for the reader as Problem 7.8.9. □

Proposition 7.5.13. *Let $f : A \to B$ and $g : B \to C$ be bijective functions. Define $h : A \to C$ to be $h = g \circ f$. Then h is invertible and $h^{-1} = f^{-1} \circ g^{-1}$.*

Proof. Left for the reader as Problem 7.8.10 □

7.5.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) Is the composition of functions **associative**? (That is, does the order of parentheses matter?) Why or why not?

(2) Is the composition of functions **commutative**? (That is, can we reverse the order?) Why or why not?

(3) Suppose $f : A \to B$ and $g : B \to A$ are functions. How do we **prove** that $g = f^{-1}$?

(4) Suppose $f : A \to B$ is a bijection. Is its inverse also a bijection?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Let O be the set of odd natural numbers and let E be the set of even natural numbers. Define a function $f : O \to E$ that is a **bijection** and prove that it is so by finding its inverse.

(2) In this problem, we want you to construct an example that shows the importance of verifying **both** compositions yield the identity function when we're trying to find the inverse of a function.

Define sets A, B and functions $f : A \to B$ and $g : B \to A$ such that

$$\forall x \in A \text{.}\ g(f(x)) = x$$

but

$$\exists y \in B \,.\, f(g(y)) \neq y$$

(**Suggestion:** You might find an example where A and B both only have one or two elements ... Or, you might find an example where $A = B = \mathbb{N}$.)

(3) Let $U = \{y \in \mathbb{R} \mid -1 < y < 1\}$ and $I = \{y \in \mathbb{R} \mid -6 < y < 12\}$.

Let $g : U \to I$ be the function defined by $\forall x \in U \,.\, g(x) = 9x + 3$.

Prove that g is a bijection by finding g^{-1}.

(4) Define the function $f : \mathbb{Z} \to \mathbb{N}$ by

$$\forall z \in \mathbb{Z} \,.\, f(z) = \begin{cases} -2z + 2 & \text{if } z \leq 0 \\ 2z - 1 & \text{if } z > 0 \end{cases}$$

Prove that f is a bijection by finding f^{-1}.

(**Hint:** Your proposed inverse function will also be piece-wise defined. Be careful about the cases that will then come up in your proof.)

(5) **Challenge:** Define $I = \{y \in \mathbb{R} \mid -1 < y < 1\}$. Find a function $f : I \to \mathbb{R}$ that is a bijection and prove that it is.

(**Hint:** You do *not* need to use any trigonometric functions. Consider using $|x|$ somewhere in your expression ...)

7.6 Cardinality

7.6.1 Motivation and Definition

One important reason for caring about bijections is that they allow us to compare the **sizes** of sets! This is a notion for which you have some intuitition. For example, it's pretty clear that the set

$$\{1, 2, 3, 4, 5\}$$

has 5 elements. It is **finite**. However, the set

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$$

is **infinite**. We also understand that $\mathbb{Z}$ is an infinite set. So are $\mathbb{Q}$ and $\mathbb{R}$. What are their sizes? Can we even compare them? How could we do so *mathematically*? What does it really mean to be an *infinite* set? Are there "different infinities"?

Bijections "Pair" Elements

Let's say there are 5 pens and 5 books on a table in front of us. But also, let's pretend that we didn't know how to **count** them. How could we verify that there are just as many pens as there are books? Instead of saying, "There are 5 pens and 5 books, and $5 = 5$", can we somehow show the *set* of books and the *set* of pens has the same *size*, without knowing what that size is?

This is where a *bijection* comes into play. We can *pair off* the pens and books one-by-one. We can line them up on the table and draw a line between them, showing a correspondence between them. In the language of sets, we are identifying a *bijection* between the set of pens and the set of books. This idea is so important, that we want to impress it upon you with a quote:

In the land of Cardinality, the Bijection is King.

Imagine our study of cardinality is a journey through the Kingdom of Cardinality. In this Kingdom, we bow down to King Bijection, for he rules all. Only he can tell us when two sets have the *same cardinality*, whether they be finite or infinite.

Moreover, we really *need* to use this set terminology, because we will see some surprising and counter-intuitive results. Using these formal definitions and concepts will allow us to be rigorous and precise. The examples and results we see might blow our minds a little bit (or a lot!), but having them rooted in concepts we've already seen and theorems we've already proven lets us actually *belive* these results, mathematically speaking!

Definitions and Notation

First, let's define what it means to be **finite**.

Definition 7.6.1. *Let S be any set. We say S is* **finite** *if and only if*

$$\exists n \in \mathbb{N} \cup \{0\} \text{ such that there exists a bijection } f : S \to [n]$$

In this case, we write $|S| = n$ to indicate that the **size** *of S is n.*

Note: The empty set $S = \varnothing$ is finite, since $[0] = \varnothing$. This is why we said $n \in \mathbb{N} \cup \{0\}$ in the definition, and not just $n \in \mathbb{N}$. The function $f : \varnothing \to \varnothing$ that is a bijection is simply the *empty relation*. (Remember that a function is a relation!)

By definition, sets of the form $[n]$ are finite. They are our standard examples of finite sets, with size $|[n]| = n$. Thus, to show that a set S has size n, we need to find a bijection between S and $[n]$. For example, consider the set $\{1, 3, 5\}$. This clearly looks like it has size 3. We can show this by exhibiting the bijection $f : \{1, 3, 5\} \to [3]$ defined by $f(1) = 1$ and $f(3) = 2$ and $f(5) = 3$.

It's interesting to think about whether a finite set could have two *different* sizes. The definition technically doesn't preclude this, but we can *prove* that the size of a finite set is unique. Think about how to do that ... We will do so after a few more essential definitions.

Definition 7.6.2. *Let S be any set. We say S is* **infinite** *if and only if S is* not finite*.*

That is, S is infinite if $\forall n \in \mathbb{N} \cup \{0\}$, every possible function $f : S \to [n]$ fails to be a bijection.

When S is infinite, we use $|S|$ to indicate the **cardinality** *of the set.*

It might seem silly to define infinite in this way—*not* finite—but it certainly reflects the intuitive dichotomy between the two concepts. A set can't be *both* finite and infinite, so rather than come up with a way to categorize both of them, let's catgeorize one and define the other to be "anything else".

Also, we do **not** write $|S| = \infty$ to indicate that a set is infinite. As we will see very shortly, there are actually **many different "levels" of infinite sets**. This might seem incredibly bizarre to you right now, but you will see what we mean. Yes, there are different "sizes" of infinite sets, and we will use $|S|$ to indicate the **cardinality** of S so that we may compare it to that of other sets. Writing $|S| = \infty$ would indicate there is only "one infinity", and this is very much incorrect.

Now, that said, we are mostly going to distinsuigh just two *types* of infinite sets, for our purposes. We are doing this to show you some striking results about the sets we are already familiar with, namely $\mathbb{N}$ and $\mathbb{Z}$ and $\mathbb{Q}$ and $\mathbb{R}$. The following definition tells us what these two types are.

Definition 7.6.3. *Let S be any set.*

We say S is **countably infinite** *if and only if there exists a bijection $f : S \to \mathbb{N}$.*

We say S is **uncountably infinite** *(or just* **uncountable***) if and only if S is infinite and every function $f : S \to \mathbb{N}$ fails to be a bijection.*

Given an infinite set S, this definition establishes two possibilities for S, based on how its cardinality $|S|$ compares with $\mathbb{N}$. We use the term **countably infinite** because it represents why we intuitively think of $\mathbb{N}$ as infinite. The set $\mathbb{N}$ has "a lot" of elements, so many so that if we tried to count them we would never finish; however, the fact that we can even try to count them in this way indicates something special. There is a 1st element of $\mathbb{N}$, and a 2nd element, and a 3rd, and ... We can't name them all in our lifetime, but we could program a magical, immortal robot to print them out one-by-one. If we thought of a natural number ahead of time, no matter how huge that number is, we know the robot will *eventually* print out that number.

Perhaps we can't do this with *all* infinite sets, though. This is what the notion of an **uncountably infinite** set is meant to convey. Such a set is infinite, so there is no correspondence with a set of the form $[n]$, but it is also *"so large"* that we cannot identify a "1st element" and a "2nd element" and a "3rd element" and ... This is what a **bijection** $f : S \to \mathbb{N}$ would convey, a way to *label* all the elements of S in a way that shows they are paired off with the natural numbers. If we *cannot* do this, then the set is uncountably infinite. Now, you might not believe that such sets exist! Don't worry, we will show you some. For now, just

be aware of the distinction between **countably** and **uncountably** infinite: the difference rests on whether a **bijection** with $\mathbb{N}$ exists.

Comparing Cardinalities

As we mentioned, when S is infinite, we use $|S|$ to **compare** the cardinality of S to that of other sets. We won't write something like $|S| = \infty$. Rather, we will write something like $|S| = |T|$ to indicate that S and T have the *same* cardinality, whatever that may be. We might also write something like $|S| < |P|$ to indicate P has a *strictly larger* cardinality than S. The following definition tells us how the comparison of cardinalities is based on functions and, specifically, different kinds of jections.

Definition 7.6.4. *Let S, T be any sets.*

- *We write $|S| = |T|$ if and only if there exists a* **bijection** *$f : S \to T$.*

  In this case, we say S has the **same cardinality** *as T.*

- *We write $|S| \leq |T|$ if and only if there exists an* **injection** *$f : S \to T$.*

  In this case, we say S has cardinality **at most** *$|T|$.*

- *We write $|S| < |T|$ if and only if $|S| \leq |T|$ and $|S| \neq |T|$.*

  In this case, we say S has a **strictly smaller** *cardinality than T.*

- *We write $|S| \geq |T|$ if and only if there exists a* **surjection** *$f : S \to T$.*

  In this case, we say S has cardinality **at least** *$|T|$.*

- *We write $|S| > |T|$ if and only if $|S| \geq |T|$ and $|S| \neq |T|$.*

  In this case, we say S has a **strictly larger** *cardinality than T.*

Let's explain the motivation behind these definitions in two different ways:

In general, $f : A \to B$ being an *injection* tells us $|A| \leq |B|$ and $g : A \to B$ being a *surjection* tells us $|A| \geq |B|$. Think about schematic diagrams for the functions f and g to see why this definition makes sense. Having an injection from $A \to B$ means we can definitely "pair" the elements of A to elements of B without overlapping, but perhaps there are "many more" elements of B left over. Likewise, having a surjection from $A \to B$ means we can definitely "cover" all of B with elements of A, but maybe we had to overlap sometimes to do this, so A could have "more" elements than B. Having both of these situations together (i.e. a *bijection* from A to B) means that A and B actually have the same cardinality: we can pair off *all* their elements. This is an intuitive explanation to motivate these definitions, mind you. These types of explanations are not rigorous proofs. But now that we have *made* these definitions, we can *use* them to prove and disprove statements! To compare cardinalities of sets—even infinite ones—we just need to find a function with an appropriate property. All of our work in the rest of this chapter will be quite helpful in our journey through the Kingdom of Cardinality.

Another way to think about these definitions is that "has the same cardinality as" is an "equivalence relation" on the "set of all sets". We have to put quotes around these phrases because, as we explained in detail in Section 3.3.5 about Russell's Paradox, there is *no such thing* as the "set of all sets". Thus, it doesn't make mathematical sense in our context to talk about an equivalence relation on that "set". In some fuzzy sense, though, this is what's going on:

- Given any set S, there is certainly a bijection with itself: the identity function, $\mathrm{Id}_S : S \to S$. This shows $|S| = |S|$, i.e. the "has the same cardinality as" relation is "reflexive".
- Suppose $|S| = |T|$, so there is a bijection $f : S \to T$. Is there a bijection $g : T \to S$, as well? Why yes, we can use $g = f^{-1}$, of course! We know that is also a bijection. This shows $|T| = |S|$ via a bijection, as well, i.e. the "has the same cardinality as" relation is "symmetric".
- Suppose $|S| = |T| = |U|$, so there are bijections $f : S \to T$ and $g : T \to U$. Does there exist a bijection $h : S \to U$, as well? Yes! The composition $g \circ f$ is also a bijection (this is something you will prove/have proven in the exercises). This shows $|S| = |U|$ via a bijection, as well, i.e. the "has the same cardinality as" relation is "transitive".

Again, this is not *exactly* what's going on, but it can really help you sort through these difficult, abstract ideas. We are establishing a way to take any two sets and compare their cardinalities using functions. All of the sets in the universe will be "partitioned" into different "classes" based on their cardinalities. What's truly amazing is what we are about to prove for you: that there are *infinitely-many cardinalities*.

Cantor's Theorem

The following result and proof are due to the German mathematician Georg Cantor from the mid- to late-1800s. By now, mathematicians have fully embraced the result and its consequences. However, at the time, this idea was so controversial that some mathematicians refused to believe him. In time, though, his work and ideas helped lead to the development of formal set theory.

The proof of this particular result is known as **Cantor's Diagonalization Argument**. We will use an argument like this later on, where we will point out why it is like a "diagonal". For now, we are more interested in the conclusion of this theorem.

Theorem 7.6.5. *Let S be any set. Then $|S| < |\mathcal{P}(S)|$.*

This says that **the power set of a set always has *strictly larger* cardinality than the set itself**. This makes sense for finite sets. You discovered already that the power set of $[n]$ has 2^n elements, i.e. $|\mathcal{P}([n])| = 2^n$. (You will prove this by induction, using results about cardinality, in Problem 7.8.30.) We see, indeed, that $n < 2^n$ for every $n \in \mathbb{N}$. However, this theorem also asserts that

this relationship holds for **infinite** sets. Wow! Immediately, this tells us that there is a whole chain of infinite sets, each one bigger than the previous one. We can just kept taking the power set of what we had before:

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < \left|\mathcal{P}\big(\mathcal{P}(\mathbb{N})\big)\right| < \left|\mathcal{P}\Big(\mathcal{P}\big(\mathcal{P}(\mathbb{N})\big)\Big)\right| < \cdots$$

Let's prove this theorem. The proof is very short and clever, so don't worry about how to *come up* with such an argument. Focus on understanding the logical flow.

Proof. Let S be any set. AFSOC $|S| \geq |\mathcal{P}(S)|$.

This means there exists a function $g : S \to \mathcal{P}(S)$ that is surjective.

Define $T = \{X \in S \mid X \notin g(X)\}$. (This makes sense because, for any $X \in S$, $g(X) \in \mathcal{P}(S)$, i.e. $g(X) \subseteq S$. Thus, either $X \in g(X)$ or $X \notin g(X)$ must hold.)

Notice $T \subseteq S$, by a set-builder notation definition. This means $T \in \mathcal{P}(S)$.

Since g is surjective, $\exists Y \in S$ such that $g(Y) = T$. Let such a Y be given.

Now, is $Y \in T$? We consider both cases:

- If $Y \in T$, then the definition of T says $Y \notin g(Y)$. However, $g(Y) = T$, so this means $Y \notin T$. This is a contradiction ⨳
- If $Y \notin T$, then the definition of T says $Y \in g(Y)$. However, $g(Y) = T$, so this means $Y \in T$. This is a contradiction ⨳

In either case, both $Y \in T$ and $Y \notin T$ hold. This is a contradiction ⨳

Therefore, there exists no such surjection from S to $\mathcal{P}(S)$, i.e. $|S| < |\mathcal{P}(S)|$. □

Look back at Exercise 4 in Section 7.4.5. Notice that we asked you to define a function from $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$, and then we asked you to prove it was **not surjective**. We didn't have to know what your function was! Since we were aware of this theorem, we knew you couldn't *possibly* have defined a surjection!

Discussion: Axioms and Definitions

We want to make an admission. We have glossed over some details about what constitutes a *definition* as opposed to a *theorem*, a result that needs proven from fundamental assumptions. By *definition* (at least, in our context) an injection and a surjection from A to B (in that direction, mind you) constitute sufficient proof of equal cardinalities, which guarantees a bijection. Likewise, an injection from A to B and one from B to A is sufficient to guarantee $|A| = |B|$, and so there must be a bijection between A and B.

It is not *totally obvious*, though, why these claims should be true. Say we have an injection from A to B and one from B to A. Does this *guarantee* a bijection between the two sets? Well, one would hope! But this isn't a proof. This result is actually known as the **Cantor-Schroeder-Bernstein Theorem**:

Theorem 7.6.6 (Cantor-Schroeder-Bernstein)**.** *Suppose A, B are any sets, and $f : A \to B$ and $g : B \to A$ are injections. Then there exists a* bijection *$h : A \to B$.*

Yes, that is a *theorem*; it is not trivial! One of the proofs is, in fact, *constructive*: it provides an algorithmic method for constructing that bijection $h : A \to B$, using the two injections, $f : A \to B$ and $g : B \to A$. For our purposes—and for time and space restrictions—there is no need to separate this out as a theorem, let alone one with a constructive proof. It is sufficient to consider injections and surjections and their consequences vis-à-vis cardinalities as *definitions*; these results "feel" intuitive and we can accept them. Just realize, though, that we are basing them on rigorous mathematical knowledge. If you are interested in learning about these subtleties and their consequences, consider taking a course or reading a book about **set theory**.

In essence, the real issue is that we pre-supposed *any* two sets, A and B, can have their cardinalities *compared* in some meaningful way, mathematically speaking. That is, for any A and B, we have pre-supposed that we can somehow declare that $|A| \leq |B|$ or $|B| \leq |A|$ makes sense (or both, perhaps, if the sets are of "equal size"). But how can we *guarantee* one such comparison, or maybe both, will always apply, for any two given sets? It's not a trivial consideration! In the context of this book, one of our axioms is that the cardinalities of any two sets we consider can be compared. In the context of the mathematical universe at large, though, this is something that needs to be proved from more fundamental assumptions.

7.6.2 Finite Sets

Before moving into the somewhat bizarre (but fascinating!) world of infinite sets, let's focus on some results about **finite** sets. These results will be easier to understand, intuitively, and will give us some good practice in working with functions and their properties to prove facts about cardinalities.

Theorems

For each of these results, we will state a theorem/proposition/lemma, and either prove it or have you help us with the proof via some exercises.

Theorem 7.6.7. *Suppose A, B are* disjoint *finite sets. Then $|A \cup B| = |A| + |B|$.*

Play around with some examples to see why this claim is True. Do you see why we need the sets to be *disjoint* for this to work? Can you prove this claim? Remember that we want to find a *bijection* between the two sets ...

Proof. Let A, B be finite sets that are disjoint.

We know $\exists a, b \in \mathbb{N} \cup \{0\}$ and there exist bijections $f : A \to [a]$ and $g : B \to [b]$.

(That is, we suppose $|A| = a$ and $|B| = b$). Let such a, b, f, g be given.

WWTS $|A \cup B| = |A| + |B| = a + b$; that is, WWTS there is a bijection $h : A \cup B \to [a + b]$.

Define the function $h : A \cup B \to [a + b]$ by

$$\forall x \in A \cup B\,. \quad h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) + a & \text{if } x \in B \end{cases}$$

Notice that h is well-defined because $A \cap B = \varnothing$, so every $x \in A \cup B$ satisfies $x \in A$ or $x \in B$ and certainly not both. Also, $1 \leq h(x) \leq a$ for every $x \in A$, and $a + 1 \leq h(x) \leq a + b$ for every $x \in B$, so $h(x) \in [a + b]$ for every x in the domain of h.

We claim that the function $H : [a + b] \to A \cup B$, defined by

$$\forall y \in [a + b]\,. \quad H(y) = \begin{cases} f^{-1}(y) & \text{if } 1 \leq y \leq a \\ g^{-1}(y - a) & \text{if } a + 1 \leq y \leq a + b \end{cases}$$

is the inverse of h. If this holds, then we have proven that h is a bijection.

Let's show that H is well-defined. Every $y \in [a + b]$ satisfies exactly one of the two inequalities given in the definition of H. Also, f and g were given to be bijections, so f^{-1} and g^{-1} are well-defined functions (that are bijections themselves, even). Furthermore, if $a + 1 \leq y \leq a + b$ then $1 \leq y - a \leq b$ so $y - a \in [b]$ (the domain of g^{-1}).

Let's show that $h \circ H = \mathrm{Id}_{[a+b]}$. Let $y \in [a + b]$ be given. We have two cases.

(1) Suppose $1 \leq y \leq a$; that is, suppose $y \in [a]$. Then,

$$(h \circ H)(y) = h(H(y)) = h\big(f^{-1}(y)\big) = f\big(f^{-1}(y)\big) = \mathrm{Id}_{[a]}(y) = y$$

where we used the fact $f^{-1}(y) \in A$.

(2) Suppose $a + 1 \leq y \leq b$; that is, suppose $y - a \in [b]$. Then,

$$\begin{aligned}(h \circ H)(y) = h(H(y)) &= h\big(g^{-1}(y - a)\big) = g\big(g^{-1}(y - a)\big) + a \\ &= \mathrm{Id}_{[b]}(y - a) + a = (y - a) + a = y\end{aligned}$$

where we used the fact that $g^{-1}(y - a) \in B$.

In either case, we find $(h \circ H)(y) = y$, and both cases are disjoint and cover all possibilities.

Next, let's show that $H \circ h = \mathrm{Id}_{A \cup B}$. Let $x \in A \cup B$ be given. We have two cases.

(1) Suppose $x \in A$. Then,

$$(H \circ h)(x) = H(h(x)) = H\big(f(x)\big) = f^{-1}\big(f(x)\big) = \mathrm{Id}_A(x) = A$$

where we used the fact that $f(x) \in [a]$.

(2) Suppose $x \in B$. Then,

$$(H \circ h)(x) = H(h(x)) = H\big(g(x) + a\big) = g^{-1}\Big(\big(g(x) + a\big) - a\Big)$$
$$= g^{-1}\big(g(x)\big) = \mathrm{Id}_B(x) = x$$

where we have used the fact that $g(x) \in [b]$ so $a + 1 \leq g(x) + a \leq a + b$.

In either case, we find $(H \circ h)(x) = x$, and both cases are disjoint and cover all possibilities.

Thus, $H = h^{-1}$, so h has an inverse. Therefore, h is a bijection.

Therefore, $|A \cup B| = \big|\,[a + b]\,\big| = a + b = |A| + |B|$. □

Corollary 7.6.8. *Suppose S, T are finite sets and $S \subseteq T$. Then, $|T - S| = |T| - |S|$.*

Proof. Define $U = T - S$. Notice that $U \cap S = \varnothing$. Apply the above theorem to U and S to get

$$|U| + |S| = |U \cup S| = |T|$$

then subtract from both sides to get $|T - S| = |U| = |T| - |S|$. □

You can use the two results above to prove the following generalization:

Proposition 7.6.9. *Suppose A, B are finite sets. Then $|A \cup B| = |A| + |B| - |A \cap B|$.*

Proof. Left for the reader as Exercise 1 in Section 7.6.5. □

Here's another corollary to the theorem above.

Corollary 7.6.10. *Suppose $A_1, A_2, \ldots, A_n$ are finite and pairwise-disjoint (remember this means any two of the sets are disjoint).*

Then $|A_1 \cup \cdots \cup A_n| = |A_1| + \cdots + |A_n|$.

Proof. Left for the reader as Exercise 2 in Section 7.6.5. □

You should also look at Problem 7.8.32 in this chapter's exercises. There, we guide you through a proof (by induction on two variables!) about the size of the *Cartesian product* of two finite sets.

7.6.3 Countably Infinite Sets

Let's move on to investigate the land of countably infinite sets. We will start by talking about a famous thought experiment, named after the mathematician David Hilbert.

The Hilbert Hotel

Let's play make-believe. This will help us get a handle on infinite weirdness.

Pretened we own a hotel. There are countably infinitely many rooms in our magical building. They are numbered as Room 1, Room 2, Room 3, ... That is to say, our rooms are *indexed* by the set of natural numbers, $\mathbb{N}$.

We want to accomodate as many people as we can (to make lots of money!) and because our hotel is so swanky and accomodating, our guests are totally willing to move to a new room whenever we ask them to. It just takes them a couple of minutes to gather their belongings and walk down the hall to a new room.

We also have a loudspeaker system that allows us to communicate a message to all of the guests at once.

- Suppose all the rooms are full. It's a very busy weekend. One guy walks into the lobby looking for a room. Can we squeeze him in? If not, why? If so, how?

  It turns out that we can! We can just shift all the guests down one room and place this new guy into Room 1.

  The catch, though, is to take advantage of our loudspeaker system. If we had to go and knock on *everyone's* door telling them to move down one room, we would *never actually finish*; we would spend all of eternity knocking on doors and delivering messages.

  Instead, we make the following announcement:

  > Attention guests: If you find yourself in Room n, please move to Room $n+1$. Thank you!

  After five minutes, the guests have all moved, and Room 1 is vacant for our new guest.

  Morally speaking, we have just verified that the set $\mathbb{N}$ and the set $\mathbb{N} \cup \{\star\}$ have the same cardinality, for any particular object $\star$. In particular, say, $|\mathbb{N}| = |\mathbb{N} \cup \{0\}|$. Our hotel has only countably many rooms, and we have accomodated one person associated with each natural number, as well as one *more* person.

- It's the next day. Our rooms are still full. Suppose a Scrabble convention with countably infinitely many people shows up. The people are all wearing nametags with natural numbers on them, so there is Person 1, Person 2, Person 3, ...

  Can we accomodate these folks? How can we assign them rooms? How do we move around the guests currently in the hotel?

  It turns out that we can! The idea is to free up an infinite set of rooms.

Again, the catch is to do this by making one blanket announcement to *all* of the guests at once, as opposed to knocking on everyone's door.

We recognize that the set of even-numbered rooms and the set of odd-numbered rooms are both infinite in size, so let's make the current guests in the hotel occupy the even-numbered rooms and assign the new guests from the convention to the odd-numbered rooms. We make the following announcement to the hotel guests via the loudspeaker:

> Attention guests: If you find yourself in Room n, please move to Room $2n$. Thank you!

Then, we make the following announcement to the convention folks waiting in the lobby:

> Attention convention-goers: If you are wearing nametag number n, please go to Room $2n - 1$. Thank you!

After five minutes, every hotel guest has moved, and after another five minutes, every convention-goer has found their room. Voilá!

Morally speaking, we have just verified that the union of two disjoint countably infinite sets is countably infinite, as well. That is, we took the set A of current hotel guests (notice A is countably infinite) and the set B of convention guests waiting for rooms (notice B is countably infinite, and notice that $A \cap B = \varnothing$) and found a bijection between $A \cup B$ and $\mathbb{N}$, where $\mathbb{N}$ represents the set of Rooms.

- Now, suppose another convention shows up. They play Scrabble in a different language, so they don't want to be associated with the other convention. How can we move folks around to get everyone a room?

  We can do the exact same thing! It's as if we were facing the same situation as before, with just a full hotel and a countably infinite set of people waiting for rooms.

- Now, suppose countably infinitely many conventions show up, each of them not wanting to be associated with the others. Oh my!

  Luckily, the hotel convention organizer has assigned every convention a natural number, and each member within a convention gets a hat with that number on it. Also, within each convention, each person is assigned a natural number, and they wear a badge with that number. Thus, each person has two forms of identification: a hat and a badge. So we have Person 1 from Convention 1, and Person 3 from Convention 7, and Person 12 from Convention 8, and so on and so forth.

  How do we rearrange all of these people in the hotel? Can we even do it? How can we do it *efficiently*?

The catch here is that we **cannot** apply the same method as the previous two cases over and over. Yes, we can squeeze in Convention 1 using that method. After that's done, we would squeeze in Convention 2. And so on. But **never** would we get to *all* of the conventions. It's the same problem we had before where knocking on every individual door would take forever to accomplish; we needed to send a message to *everyone at once*. Likewise, here, we need to send a message to all of the hotel guests, and then a message to all of the convention-goers waiting outside the door. It needs to be a general "formula" about which room to go to.

If it helps, think of this from the other side of the situation. Pretend you are in Convention x and you are Person y in that convention. You are eagerly awaiting a comfortable bed to sleep in for the night. You want to know *exactly* what room to go to. ASAP. You don't want to wait around and see all of the conventions ahead of you given rooms, one by one. You want to *all* go in at once and find your corresponding rooms.

Here's one way to do it. Let's take advantage of the structure of the *prime numbers*. We know there are countably infinitely many primes, and that for any two *different* primes p and q (i.e. $p \neq q$), it is true that $p^k \neq q^k$ for any natural number k. With this in mind, we see that assigning individual conventions to the rooms that are powers of a corresponding prime number, we can ensure that no two (potential) guests get assigned to the same room. We make the following announcement to our current hotel guests:

> Attention guests: If you find yourself in Room n, please move to Room 2^n. Thank you!

We then make the following announcement to the conventions waiting outside the door:

> Attention convention-goers:
>
> If you are Person number k from Convention number 1, please go to Room 3^k.
>
> If you are Person number k from Convention number 2, please go to Room 5^k.
>
> If you are Person number k from Convention number 3, please go to Room 7^k.
>
> In general, if you are Person number k from Convention number n, please go to the Room numbered by the $(n+1)$-th prime number raised to the k-th power.
>
> Thank you!

(Note: We are assuming that all of our guests and potential guests are math genii, and they can quickly figure out what the $(n+1)$-th prime

number is and raise it to the k-th power. Otherwise, we wouldn't want them to stay at our luxurious, mathematical hotel in the first place!)

Notice that this guarantees *everyone* has a room all to themself. Nobody has to share a room. However, it *does* leave many rooms *empty*. Who is in Room 1? Room 6? Room 18? In general, can you characterize the set of rooms that will be empty?

How could we have been more "efficient" about this? Is there a certain announcement we could make so that *all* the rooms are filled?

Morally speaking, we just verified that $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$ have the same cardinality. We had countably infinitely many conventions with countably infinitely many people in each, so every person we wanted to accomodate corresponds to an *ordered pair of natural numbers*, where the first coordinate is their Person number and the second coordinate is their Convention number. Since we were able to match this set of people with the set of rooms (which corresponds to $\mathbb{N}$), then we showed $\mathbb{N} \times \mathbb{N}$ is countable. (Note: We actually "overdid" it and found a way to embed the set $\mathbb{N} \times \mathbb{N}$ in a *strict subset* of $\mathbb{N}$!)

This hopefully gives you a flavor for how to think about countable infinities. One important point to keep in mind is that **infinity** is a **cardinality**, not a **number**, in our context here. It's not as if the natural numbers "keep going" and there's some magical number ∞ lying out there past them all. Here, we refer to countably *infinite* as a **cardinality**; it represents how "big" something is. It's more like a *magnitude* than a *position*.

Examples

Let's take some of the ideas conveyed by the **Hilbert Hotel** examples and express them more formally. We'll make use of injections and surjections and bijections. (Oh my!) The following result will be helpful as we go along, so let's prove it now.

Lemma 7.6.11. *Let S, T be any sets. Suppose $S \subseteq T$. Then $|S| \leq |T|$.*

Proof. Define the "identity function" $f : S \to T$, given by $\forall x \in S \text{ . } f(x) = x$.

Since $S \subseteq T$, this is a well-defined function.

(Note: We couldn't technically define this as the usual identity function Id_S, because the domain and codomain might not be equal sets; in essence, f does the same action as Id_A but has a different codomain).

Notice that f is injective!

(Note: It's not necessarily bijective, because it might be that $S \neq T$.)

Since f is injective, this tells us that $|A| \leq |B|$. □

You might be wondering why we can't conclude $|A| < |B|$ here. Why is it "$\leq$" instead? Certainly, $\{1,2\} \subseteq \{1,2,3\}$ and $|\{1,2\}| = 2 < 3 = |\{1,2,3\}$. This is true for **finite** sets, but as we shall see in this section, there are **infinite** sets that have *strict* subsets of equal cardinality!

Example 7.6.12. $\mathbb{Z}$ **is countably infinite:**

We know $\mathbb{N}$ is countably infinite by *definition*. The identify function $\mathrm{Id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ is obviously a bijection, so $\mathbb{N}$ is countable.

In this example, we will prove that $\mathbb{Z}$ is countably infinite! To accomplish this, we need to find a bijection $f : \mathbb{Z} \to \mathbb{N}$. We will state one here and then prove it is a bijection by finding its *inverse*. Before reading on, try to find a bijection on your own! Maybe you'll come up with a different function than ours! If you need a hint for coming up with one, think about this: to prove an infinite set is *countably* infinite, we want to find a way to start *listing* the elements one by one. Try to find a pattern that identifies the "1st" integer, and then the "2nd, and then the "3rd", ...

Let's define a function $f : \mathbb{Z} \to \mathbb{N}$ and then prove it is a bijection by identifying f^{-1}.

Explicit bijection: We choose to define $f : \mathbb{Z} \to \mathbb{N}$ by setting

$$\forall z \in \mathbb{Z}.\ f(z) = \begin{cases} -2z + 2 & \text{if } z \leq 0 \\ 2z - 1 & \text{if } z > 0 \end{cases}$$

We chose this function because it "pairs off' the integers with the natural numbers like this:

$$\begin{array}{cccccccccc} \ldots, & -3, & -2, & -1, & 0, & 1, & 2, & 3, & \ldots \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \ldots, & 8, & 6, & 4, & 2, & 1, & 3, & 5, & \ldots \end{array}$$

(That is, we are pairing off the even natural with the non-positive integers, and the odd natural with the positive integers. Looking at this correspondence, we can see how to "reverse" it. This is how we will find f's inverse.)

Next, Define $F : \mathbb{N} \to \mathbb{Z}$ by

$$F(n) = \begin{cases} -\frac{n}{2} + 1 & \text{if } n \text{ is even} \\ \frac{n+1}{2} & \text{if } n \text{ is odd} \end{cases}$$

Let's show $F = f^{-1}$. Let $z \in \mathbb{Z}$ be given. We have two cases.

- Suppose $z \geq 1$. Then $f(z) = 2z - 1$. Notice that $2z - 1 \in \mathbb{N}$ and $2z - 1$ is odd. This means

$$(F \circ f)(z) = F(f(z)) = F(2z-1) = \frac{(2z-1)+1}{2} = \frac{2z}{2} = z$$

- Suppose $z \leq 0$. Then $f(z) = -2z+2$. Notice that $-2z \geq 0$ so $-2z+2 \geq 2$ so $-2z+2 \in \mathbb{N}$. Also, $-2z+2$ is even. This means

$$\begin{aligned}(F \circ f)(z) = F(f(z)) = F(-2z+2) &= -\frac{-2z+2}{2} + 1 \\ &= -(-z+1) + 1 = (z-1) + 1 = z\end{aligned}$$

In either case, $(F \circ f)(z) = z$. This shows $F \circ f = \mathrm{Id}_{\mathbb{Z}}$.

Next, let $n \in \mathbb{N}$. We have two cases.

- Suppose n is even. Then $F(n) = -\frac{n}{2} + 1$. Notice that $\frac{n}{2} \geq 1$ and so $-\frac{n}{2} \leq -1 + 1 = 0$. This means

$$\begin{aligned}(f \circ F)(n) = f(F(n)) = f\left(-\frac{n}{2}+1\right) &= -2\left(-\frac{n}{2}+1\right) + 2 \\ &= \left(\frac{2n}{2} - 2\right) + 2 = n\end{aligned}$$

- Suppose n is odd. Then $F(n) = \frac{n+1}{2}$. Notice that $n+1 \geq 2$ and so $\frac{n+1}{2} \geq 1$. This means

$$\begin{aligned}(f \circ F)(n) = f(F(n)) = f\left(\frac{n+1}{2}\right) &= 2\left(\frac{n+1}{2}\right) - 1 = \frac{2n+2}{2} - 1 \\ &= (n+1) - 1 = n\end{aligned}$$

In either case, $(f \circ F)(n) = n$. This shows $f \circ F = \mathrm{Id}_{\mathbb{N}}$. Therefore, $F = f^{-1}$. □

This shows that $\mathbb{Z}$ and $\mathbb{N}$ have the same cardinality, that $|\mathbb{Z}| = |\mathbb{N}|$. You might feel like there are "twice as many" integers as naturals, but this is where your intution fails. We can *pair up* the elements of these two sets one-by-one, so they must be of the same size! This is an example that shows you why the conclusion of Lemma 7.6.11 is the best it can be. Here, $\mathbb{N} \subset \mathbb{Z}$ (a *strict* subset) and yet $|\mathbb{N}| = |\mathbb{Z}|$. This can only happen when have infinite (not finite) sets, and here is one such example.

(Later in this section, we will in fact *prove* that this is an equivalent way of characterizing when a set is infinite: whether or not we can find a bijection between the set and a *strict* subset of itself.)

Example 7.6.13. $\mathbb{N} \times \mathbb{N}$ **is countably infinite:**

With the **Hilbert Hotel** discussion in the previous section, we essentially argued for the fact that $\mathbb{N} \times \mathbb{N}$ has the same cardinality as $\mathbb{N}$. When we had infintely-countably-many conventions, each with infinitely-countably-many people in them, we could *still* fit them all into our hotel with infinitely-countably-many rooms! That was more of an intuitive discussion, though, so let's formally prove this fact here. We will find an explicit *bijection* between the two sets.

Rather than finding its inverse, though, we will prove it is surjective, and ask for your help in showing that it is injective.

Explicit bijection: Define $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by setting

$$\forall (x, y) \in \mathbb{N} \times \mathbb{N}. \quad f(x, y) = 2^{x-1}(2y - 1)$$

In proving that f is a bijection, we will be proving this fact:

> Every natural number can be written *uniquely* as a power of 2 times an odd natural number.

Look at the function we defined. It takes a pair of natural numbers and outputs a power of 2 times an odd natural number. Proving this is a bijection shows that it never outputs the same natural twice (injectivity) and every natural number is an output of some pair (surjectivity). You might try playing around with the function, plugging in some values and seeing what happens. Also, you might try working "backwards", trying to figure out what f^{-1} might possibly do. For instance, take your favorite $n \in \mathbb{N}$. Can you express it as a power of 2 times an odd? If n is odd, this is quite easy, since $2^0 = 1$. For instance,

$$11 = 1 \cdot 11 = 2^0 \cdot (2 \cdot 6 - 1) = f(1, 6)$$

(Notice that we had to use $x - 1$ and $2y - 1$ in the definition of f because we are working with $\mathbb{N}$, and $0 \notin \mathbb{N}$.)

If n is even, we can just divide by 2 iteratively until we can't anymore; what's left must be an odd number. For instance:

$$40 = 2 \cdot 20 = 4 \cdot 10 = 8 \cdot 5 = 2^3 \cdot (2 \cdot 3 - 1) = f(4, 3)$$

and

$$32 = 2 \cdot 16 = 2^2 \cdot 8 = 2^3 \cdot 4 = 2^4 \cdot 2 = 2^5 = 2^5 \cdot (2 \cdot 1 - 1) = f(6, 1)$$

This observation is crucial in proving that f is surjective:

f is surjective: We claim $\forall n \in \mathbb{N}. \; n \in \mathrm{Im}_f(\mathbb{N} \times \mathbb{N})$. We prove this by a "minimal criminal" argument.

BC: Notice that $f(1, 1) = 2^0 \cdot 1 = 1$. Thus, $1 \in \mathrm{Im}_f(\mathbb{N} \times \mathbb{N})$.

IH: Suppose we have $n \in \mathbb{N} - \{1\}$ that has no such representation as a power of 2 times an odd, i.e. suppose $n \notin \mathrm{Im}_f(\mathbb{N} \times \mathbb{N})$.

IS: We have two cases:

- If n is odd, then ... well, $n \cdot 2^0 = n \cdot 1 = n$ is such a representation. That is, we know $\frac{n+1}{2} \in \mathbb{N}$ and we see that

  $$f\left(1, \frac{n+1}{2}\right) = 2^0 \cdot \left(2 \cdot \frac{n+1}{2} - 1\right) = 1 \cdot (n + 1 - 1) = n$$

  so $n \in \mathrm{Im}_f(\mathbb{N} \times \mathbb{N})$. This contradicts our assumption that $n \notin \mathrm{Im}_f(\mathbb{N} \times \mathbb{N})$ so this case is not valid.

- If n is even, then, consider $\frac{n}{2}$. AFSOC we have a representation of $\frac{n}{2}$ as a power of 2 times an odd, i.e. suppose $\frac{n}{2} \in \text{Im}_f(\mathbb{N} \times \mathbb{N})$. This means $\exists (x, y) \in \mathbb{N} \times \mathbb{N}$**.** $f(x, y) = \frac{n}{2}$. Let such (x, y) be given. Consider, then, $f(x + 1, y)$ (which is valid since $x + 1 \in \mathbb{N}$, as well). We see that

$$f(x+1, y) = 2^{x+1} \cdot (2y - 1) = 2 \cdot \big(2^x \cdot (2y - 1)\big) = 2 \cdot f(x, y) = 2 \cdot \frac{n}{2} = n$$

  This shows we would have such a representation for n; i.e., in fact, $n \in \text{Im}_f(\mathbb{N} \times \mathbb{N})$. Again, this contradicts our assumption that $n \notin \text{Im}_f(\mathbb{N} \times \mathbb{N})$.

  Thus, $\frac{n}{2}$ *also* has no such representation, i.e. $\frac{n}{2} \notin \text{Im}_f(\mathbb{N} \times \mathbb{N})$.

We have shown that, supposing n is a counterexample to the claim, $\frac{n}{2}$ is a *smaller* counterexample to the claim. By a "minimal criminal" argument (since we proved our base case), we conclude that the claim holds for every $n \in \mathbb{N}$. This shows f is surjective. □

(Note: You might want to look back at Section 5.5.1 to refresh your memory about how "minimal criminal" arguments work.)

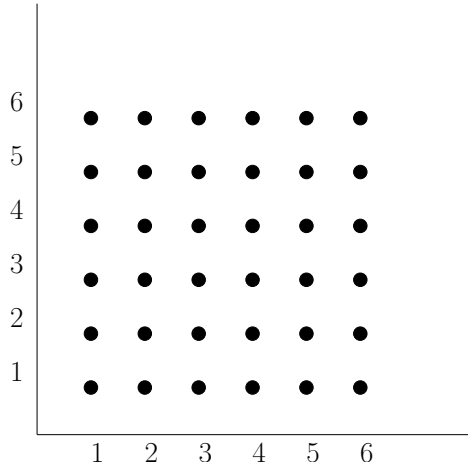
f **is injective:** You prove this! See Exercise 7.8.21.

Together, we have proven that f is a bijection, and so $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. That is, $\mathbb{N} \times \mathbb{N}$, the set of all ordered *pairs* of natural numbers, is countably infinite. Does this surprise you at all? Does it seem counter-intuitive? What do you think might be true about the set $\mathbb{N}^3$ of all ordered *triplets* of natural numbers? What do you think would happen if we took $\mathbb{N} \times \mathbb{N} \times \mathbb{N} \cdots$? Think about these ideas. Discuss them with your classmates, and try to prove something!

Example 7.6.14. $\mathbb{N} \times \mathbb{N}$ **as a lattice:**

Before moving on to another example, let's show you one more way of thinking about why $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. This will be an intuitive explanation, more like a description of how to define a bijection between the sets without actually making the definition. However, it's a common argument and is well worth seeing.

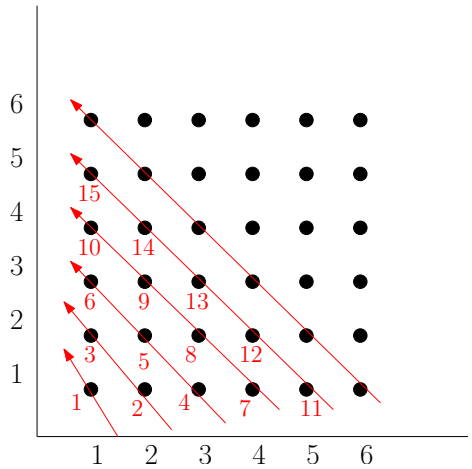
The idea is to think of $\mathbb{N} \times \mathbb{N}$ as a *lattice* of points, like so:



To show that this infinite grid of points is **countably** infinite, we can describe a path that traverses *all* of the points (surjectivity!) *exactly* once (injectivity!) and is indexed by the natural numbers (countably infinite!). That is, we can just describe a way to traverse the whole grid in a series of steps; there will be a "1st point" and a "2nd point" and so on.

The key observation to make is that the "northwestern" diagonals of this grid are all **finite**. Start from the point $(5, 1)$, for instance, and move upwards and leftwards, diagonally. You will traverse over $(4, 2)$ and $(3, 3)$ and $(2, 4)$ and $(1, 5)$, and then reach the boundary of the grid. This is true no matter *where* you start along the bottom row of lattice points.

Let's use this fact to *label* each lattice points with a natural number based on (a) which diagonal it lies on, and (b) where it lies along that diagonal. We'll treat the diagonal starting at $(1, 1)$ as the 1st diagonal, the one starting at $(2, 1)$ as the 2nd, and so on. This gives us the following labels:



We can see that every point in the lattice will lie on exactly one such diagonal. Furthermore, there are countably-infinitely-many such diagonals (they are indexed by $\mathbb{N}$) and there are only *finitely*-many points on each diagonal. This means (as we will prove below) that the collection of *all* the points on the diagonals is countably infinite.

You ought to try *formalizing* this argument by writing down a function that achieves the labeling we've demonstrated. Or, you could at least work with a similar one that also works, i.e. you could move southeastwards instead, or reverse the direction of alternate diagonals ...

Example 7.6.15. $\mathbb{Q}$ **is countably infinite:**

This result is one of the more striking examples of our intuition *failing* with infinite sets and their cardinalities. Think about the elements of $\mathbb{Q}$ as laid out on the real number line. They're everywhere! In fact, look at Exercise 4.11.26; there, you proved that the rationals are **dense**, and it is also true that they are dense *in* $\mathbb{R}$ (i.e. between any two distinct real numbers lies a rational number). Furthermore, the set of rationals *seems* so much larger than $\mathbb{Z}$: between 0 and 1 alone, there lies infinitely many rational numbers! For these reasons, you might believe that $\mathbb{Q}$ is uncountably infinite, but this is False.

In this example, we will present several arguments for this fact, especially because we realize it is so strange and striking.

(1) **Intuitive argument:**

Consider the following "representation" of $\mathbb{Q}$ as a union of sets:

$$\mathbb{Q} \text{ “=” } \text{“}\mathbb{N} \times \mathbb{N}\text{”} \cup \text{“}-\text{”}(\mathbb{N} \times \mathbb{N}) \cup \{0\}$$

In some sense, $\mathbb{N} \times \mathbb{N}$ corresponds to all the positive rationals. To see why, just consider the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}_+$ defined by $f(x, y) = \frac{x}{y}$. We definitely output all positive rationals (so f is a surjection), but $\frac{4}{2} = \frac{2}{1}$ so this is not an injection. At least, this shows $|\mathbb{N} \times \mathbb{N}| \geq |\mathbb{Q}|$ because f is a surjection. Since $\mathbb{N} \times \mathbb{N}$ is countably infinite, and we certainly expect $\mathbb{Q}$ to be infinite, this shows the positive rationals are countably infinite.

The set of negative rationals—let's call it $\mathbb{Q}_-$—must have the same cardinality as the set of positive rationals—let's call it $\mathbb{Q}_+$. There is a clear bijection between them: define $g : \mathbb{Q}_+ \to \mathbb{Q}_-$ by setting $\forall q \in \mathbb{Q}_+ \,.\, g(q) = -q$.

All this leaves out is $0 \in \mathbb{Q}$. The union of two countably infinite sets is also countably infinite (as we will prove below), and adding on one more element won't change that. Thus, $\mathbb{Q}$ is countably infinite.

Mind you, this is quite "hand-wavey". All of the "scare quotes" in the "equation" above mean you should take this as just a heuristic argument, and not a proof. However, there are ways to make all of these arguments formal. Try working on this on your own!

(2) **Listing $\mathbb{Q}$:**

Consider writing a computer program to print out all the positive rational numbers in a list. What algorithm would you use? As long as you can guarantee that your program will "eventually" succeed and print them all, then you have shown $\mathbb{Q}$ can be enumerated one-by-one, so it must be countably infinite. (Remember, this is why we use $\mathbb{N}$ as the *canonical* countably infinite set: we can enumerate its elements one-by-one, we can *count* them.)

Here's one way that we might write such a program: Follow the same "path through the lattice" argument that we used with $\mathbb{N} \times \mathbb{N}$ in the previous example. This time, though, just "skip over" any rational you have already printed.

That is, we would print the pair $(1,1) \leftrightarrow 1$ and then $(2,1) \leftrightarrow 2$ and then $(1,2) \leftrightarrow \frac{1}{2}$ and then $(3,1) \leftrightarrow 3$ and then ...

Aha! We have to omit writing $(2,2) \leftrightarrow 1$. How did we know that? We *see* that we already printed 1. How did we know that? We just looked over the list of rationals we had already printed and checked to see if what we were about to print has already appeared. If so, we move on; if not, we print it and then move on.

In terms of the enumeration process, this just means that for every point in the lattice we pass through, we have to check *finitely-many* things; namely, we have to look over the *finitely-large* set of rationals we have already printed. This means the printing process at any individual step will take "a little longer" but not *infinitely-longer*. Thus, our program will eventually print out every rational number; no matter which one you have in mind, we will get to it in finite time.

(3) $\mathbb{Q}$ **is *at most* countably infinite:**

Here's another argument about $\mathbb{Q}$ being countable. (If this feels like overkill, that's fine, just move on. We just know that this is a surprising result and having a few ways of thinking about it might help!)

Consider this: We can definitely agree a priori that $|\mathbb{Q}| \geq |\mathbb{N}|$. This follows from the fact that $\mathbb{Q} \supseteq \mathbb{N}$. Now, the only question is whether or not these cardinalities are *equal*. To reach that conclusion, we would need to find either (a) an injection from $\mathbb{Q}$ to a countable set, or (b) a surjection from a countable set to $\mathbb{Q}$.

We will prove below that that $\mathbb{Z} \times \mathbb{N}$ is countable. (That is, we will prove generally that the Cartesian product of any two countably infinite sets is also countably infinite.) We can then define the function $f : \mathbb{Z} \times \mathbb{N} \to \mathbb{Q}$ by

$$\forall (z,n) \in \mathbb{Z} \times \mathbb{N}. \quad f(z,n) = \frac{z}{n}$$

This is a surjection onto $\mathbb{Q}$. It is definitely not injective (why not?) but we don't care. It shows that $|\mathbb{Z} \times \mathbb{N}| = |\mathbb{Q}|$. Once we have proven that $|\mathbb{Z} \times \mathbb{N}| = |\mathbb{N}|$, this will have shown that $|\mathbb{N}| = |\mathbb{Q}|$.

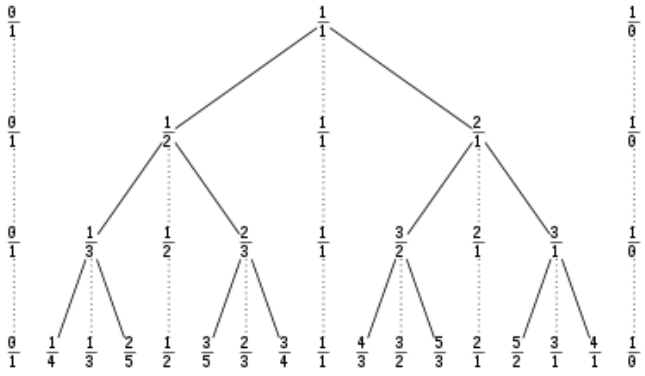
(4) **Stern-Brocot Tree:**

There are other visual representations of $\mathbb{Q}$, too! The **Stern-Brocot Tree** is particularly enlightening. This idea was, in fact, first introduced and developed by a French watchmaker named Achille Brocot who was looking for ways to approximate the measurements of gears he needed to make while building watches. Around the same time (the 1850s and 1860s), the German mathematician Moritz Stern developed the idea. It's amazing to think that a non-mathematician would *indpendently* develop this fascinating idea to solve a real-world problem he was facing!

(Do not worry too much about the terminology of *graphs* and *trees* here. We will not have occasion to talk about these much further, and are only introducing this as a helpful way to represent $\mathbb{Q}$ and demonstrate it is countably infinite.)

The **root** of this tree is 1. (This is the number at the very top of the diagram.) The **parent-child relation** (the way to generate what lies *below* a point in the tree) is defined in terms of continued fractions. (We won't describe here what that means; instead, we describe below how to *construct* the tree.)

What happens with this setup is that any path through the tree from the root to another node yields a *sequence* of rational numbers that are better and better approximations to the ultimate node; furthermore, each successive rational in the sequence has a larger denominator than the previous one. This is the property that motivated Monsieur Brocot. He needed to determine how big to make two gears inside a watch so that the *ratio* of their sizes was very close to a particular number. By working downwards through this tree, he could find better approximate ratios to the number he needed! Pretty cool, right?



To actually *construct* the tree, we find *mediants*. Given two rationals $\frac{a}{b}$ and $\frac{c}{d}$, the mediant of those two is defined as $\frac{a+b}{c+d}$. (Notice that this is a special

object, the mediant; it is *not* the correct way to add two fractions!)

Each level of the tree consists of all the mediants made from consecutive pairs of rationals in the level above; we don't "count" the directly vertical elements; they are just carried over for ease of reading and construction. Also, notice that the fractions $\frac{0}{1}$ and $\frac{1}{0}$ (which is undefined, even!) are included in the outside columns to help generate the elements on the outside of each level.

(Play around with the properties of this tree, and read more about. It is an interesting mathematical object!)

We won't prove here that this tree contains *all* the rational numbers, but we think you can see why this is believable. Also, we think you can see why the set of all nodes in this tree is **countably** infinite. Each level has only finitely many nodes, and there are countably-infinitely-many levels.

Theorems

Now we know that three of our standard sets of numbers—$\mathbb{N}$ and $\mathbb{Z}$ and $\mathbb{Q}$—are all countably infinite, as well as the set $\mathbb{N} \times \mathbb{N}$. With the following theorems, we will show you some ways to generate more countably infinite sets from existing ones.

Let's get you warmed up with one helpful result. It says that we can take a countably infinite set and "tack on" finitely-many extra elements, and this keeps the result contably infinite, as well.

Lemma 7.6.16. *If A is countably infinite and B is finite and $A \cap B = \varnothing$, then $A \cup B$ is countably infinite.*

Proof. Left for the reader as Exercise 7.8.19.

Hint: Try using a similar idea to our proof of Theorem 7.6.7. □

Remark 7.6.17. Note: The assumption that $A \cap B = \varnothing$ is not *essential* in this Lemma, but it makes the proof easier.

When $A \cap B \neq \varnothing$, we can apply the result just proven to the set $A - B$ (which is countably infinite) and the set $B - A$ (which is finite) to get the countably infinite set $(A - B) \cup (B - A)$ (since they're disjoint). We can then apply the above result again to that set—$(A-B)\cup(B-A)$—and $A \cap B$ to get the countably infinite set

$$A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$$

The next result says that this works with A, B both countably infinite, as well.

Lemma 7.6.18. *If A and B are countably infinite and $A \cap B = \varnothing$, then $A \cup B$ is countably infinite.*

Proof. Since A and B are countably infinite, there exist bijections $f : A \to \mathbb{N}$ and $g : B \to \mathbb{N}$. Let such functions be given. We will use them to find a bijection $h : A \cup B \to \mathbb{N}$.

First, define the function $p : \mathbb{N} \to \mathbb{Z} - \mathbb{N}$ by setting $\forall n \in \mathbb{N}.\ p(n) = -n + 1$.

This is a bijection because $p^{-1} : \mathbb{Z} - \mathbb{N} \to \mathbb{N}$ is given by $p^{-1}(z) = -z + 1$. (Check this for yourself!)

Since p and g are bijections, we know $p \circ g : B \to \mathbb{Z} - \mathbb{N}$ is a bijection, as well.

Next, we define the piece-wise function $q : A \cup B \to \mathbb{Z}$ by setting

$$\forall x \in A \cup B. \qquad q(x) = \begin{cases} f(x) & \text{if } x \in A \\ p(g(x)) & \text{if } x \in B \end{cases}$$

This is well-defined because $A \cap B = \varnothing$. Furthermore, this is a bijection because it is a bijection on each of the pieces is a bijection. (Again, check this for yourself to make sure it makes sense. Also, see Exercise 7.8.31 which proves this, in generality.)

From previous work, we know how to find a bijection $r : \mathbb{Z} \to \mathbb{N}$. (Remember how we did that? Look back at Example 7.6.12!)

Finally, define $h : A \cup B \to \mathbb{N}$ by $h = r \circ q$. This is a composition of bijections, so it is a bijection. This proves $|A \cup B| = |\mathbb{N}|$, i.e. $A \cup B$ is countably infinite. □

The next corollary says that we did not, in fact, *need* to assume that $A \cap B = \varnothing$. It made the proof easier. We will ask you to prove this corollary.

Corollary 7.6.19. *If A and B are countably infinite, then $A \cup B$ is countably infinite.*

Proof. Left for the reader as Exercise 7.8.20.

(**Hint:** Apply Lemma 7.6.18 to appropriately-chosen sets. . .) □

This proves several cases about finding the *union* of sets. Let's prove a result about taking a *Cartesian product*.

Theorem 7.6.20. *If A and B are countably infinite, then $A \times B$ is countably infinite.*

This one is actually easy to prove, but only because we've already proven a result about a canonical set that is a Cartesian product and is countably infinite, itself. Look at how we use $\mathbb{N} \times \mathbb{N}$ in the proof:

Proof. Suppose A, B are countably infinite. Then there exist bijections $f : A \to \mathbb{N}$ and $g : B \to \mathbb{N}$. Let such functions be given.

Define the function $h : A \times B \to \mathbb{N} \times \mathbb{N}$ by

$$\forall (x, y) \in A \times B. \quad h(x, y) = \big(f(x), g(y)\big)$$

We claim this is a bijection. Since f, g are invertible, we claim that $H : \mathbb{N}\times\mathbb{N} \to A \times B$ given by

$$\forall (k,\ell) \in \mathbb{N}\times\mathbb{N}. \quad H(k,\ell) = \big(f^{-1}(k), g^{-1}(\ell)\big)$$

satisfies $H = h^{-1}$.

To see why, notice that

$$\begin{aligned}\forall (x,y) \in A\times B. \quad (H\circ h)(x,y) &= H(h(x,y)) = H\big(f(x),g(y)\big)\\ &= \big(f^{-1}(f(x)), g^{-1}(g(y))\big) = (x,y)\end{aligned}$$

and

$$\begin{aligned}\forall (k,\ell) \in \mathbb{N}\times\mathbb{N}. \quad (h\circ H)(k,\ell) &= h(H(k,\ell)) = h\big(f^{-1}(k), g^{-1}(\ell)\big)\\ &= \big(f(f^{-1}(k)), g(g^{-1}(\ell))\big) = (k,\ell)\end{aligned}$$

so $H \circ h = \mathrm{Id}_{A\times B}$ and $h \circ H = \mathrm{Id}_{\mathbb{N}\times\mathbb{N}}$. This shows $H = h^{-1}$.

Therfore, h is a bijection, and so $|A \times B| = |\mathbb{N}\times\mathbb{N}| = |\mathbb{N}|$. □

By applying induction to the two previous results, we can prove the following:

Corollary 7.6.21. *Suppose $A_1, \ldots, A_n$ are countable (where $n \in \mathbb{N}$, so we only have finitely many sets).*

Then $A_1 \cup \cdots \cup A_n$ and $A_1 \times \cdots \times A_n$ are countably infinite.

Proof. Left for the reader as Exercise 7.8.22 □

A Countable Union of Countable Sets is Countable

You might wonder now what happens when we take a union or product of a *countably-infinite*number of sets, *each* of which is countably infinite ... Let's tackle the union case here. This result is so fundamental and important, that we've even reiterated it in the section title here!

Theorem 7.6.22. *Suppose we have, for each $n \in \mathbb{N}$, a countably infinite set A_n. Then the set*

$$A = \bigcup_{n\in\mathbb{N}} A_n = A_1 \cup A_2 \cup A_3 \cup \cdots$$

is also countably *infinite.*

We will prove this in the case that the sets are **pairwise-disjoint**, and leave the rest of the details to you.

Proof. Suppose we have, for each $n \in \mathbb{N}$, a countably infinite set A_n. Furthermore, suppose $\forall i, j \in \mathbb{N}\,.\; i \neq j \implies A_i \cap A_j = \varnothing$. Define

$$A = \bigcup_{n \in \mathbb{N}} A_n$$

We claim A is countably infinite.

Since each A_n is countably infinite, we know there exists a bijection $f_n : A_n \to \mathbb{N}$, for every $n \in \mathbb{N}$. This lets us "number" the elements of every set A_n, based on what the bijections f_n do. Furthermore, we have a number on the A_n sets (they are indexed by $\mathbb{N}$). In essence, then, we have a "numbering" of the elements of A that corresponds to $\mathbb{N} \times \mathbb{N}$. Let's formally define this correspondence.

Let's define a function $F : A \to \mathbb{N} \times \mathbb{N}$. Given any $x \in A$, we know $\exists n \in \mathbb{N}\,.\; x \in A_n$ and that this n is *unique*. (This follows because the given sets were pairwise-disjoint). Set $F(x) = \big(n, f_n(x)\big)$.

We claim that F is a bijection. To see why, consider the function $G : \mathbb{N} \times \mathbb{N} \to A$ defined by

$$\forall (a, b) \in \mathbb{N} \times \mathbb{N}\,.\; G(a, b) = f_a^{-1}(b)$$

That is, G uses the first coordinate a to identify the set A_a, and then uses the function f_a to identify the element of A_a that produced $b \in \mathbb{N}$ as an output.

(We will leave it to the reader to verify that, indeed, $G = F^{-1}$.)

This shows that $|A| = |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$, so A is countably infinite.

In the case where the A_n sets are *not* necessarily pairwise-disjoint ... we leave this as Exercise 7.8.37. □

Corollary 7.6.23. *Suppose we have, for every $n \in \mathbb{N}$, a* **finite** *set A_n. Furthermore, suppose that these sets are pairwise-disjoint. Define*

$$A = \bigcup_{n \in \mathbb{N}} A_n$$

Then A is countably infinite.

Proof. Left for the reader as Exercise 7.8.36 □

This result is very powerful. Let's see it applied to two examples.

Example 7.6.24. **The set of all powers of primes:**

Recall the Hilbert Hotel discussion, where we accomodated infinitely many conventiones of people that were each infinitely large. We sent people to rooms corresponding to powers of primes. For every $n \in \mathbb{N}$, define p_n to be the n-th prime number. Then, for every $n \in \mathbb{N}$, define

$$A_n = \{p_n^k \mid k \in \mathbb{N}\}$$

which is the set of all powers of the n-th prime. The theorem above says that

$$\bigcup_{n\in\mathbb{N}} A_n = \{\text{all powers of primes}\}$$

is countably infinite, as well. Indeed, we should have expected that because that union is just a subset of the natural numbers, which is countably infinite itself!

Example 7.6.25. **The set of all finite binary strings:**

A binary string is defined to be an ordered list of 0s and 1s. A **finite binary string** is one that is of finite length.

For example, the following are all finite binary strings:

$$0, \quad 1, \quad 101010, \quad 10000000000000000001$$

For every $n \in \mathbb{N}$, let's define F_n to be the set of all binary strings of length n. For instance,

$$\begin{aligned} F_1 &= \{\, 0\ ,\ 1\,\} \\ F_2 &= \{\, 00\ ,\ 01\ ,\ 10\ ,\ 11\,\} \\ F_3 &= \{\, 000\ ,\ 001\ ,\ 010\ ,\ 100\ ,\ 011\ ,\ 101\ ,\ 110\ ,\ 111\,\} \end{aligned}$$

and so on. (Notice that $|F_n| = 2^n$. Try to prove that!) Then, define the set of all finite binary strings by

$$F = \bigcup_{n\in\mathbb{N}} F_n$$

An element of F must have come from some set in the big union; this means that an arbirtrary element $x \in F$ is some binary string with *some* finite length. That length could be a huuuuuuuge number, but it is finite. (This points out the distinction between allowing something to be "arbitrarily large (but finite)" and allowing something to be "infinite".)

The point of this example is that F is countably infinite, according to the theorem above! (Well, it follows from the corollary stated right after, actually.) Contrast this with the set S of all *infinite* binary strings, which is—as we will prove shortly—uncountably infinite. We will use these sets of binary strings fairly often as examples!

Passing Off To A "Limit"

We proved above that if A and B are countably infinite, then so are $A \cup B$ and $A \times B$. We also encouraged you to prove (by induction on the number of sets in the union/product) that

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i\in[n]} A_i \qquad \text{and} \qquad \prod_{i\in[n]} A_i = A_1 \times A_2 \times \cdots \times A_n$$

are both countably infinite, as well, for any $n \in \mathbb{N}$.

What do these results tell us, if anything, about

$$A_1 \cup A_2 \cup A_3 \cup \cdots \quad = \quad \bigcup_{k \in \mathbb{N}} A_k$$

and

$$A_1 \times A_2 \times A_3 \cdots \quad = \quad \prod_{k \in \mathbb{N}} A_k$$

That is, what happens when we try to "jump to the limit" from having a *finite* union/product (of arbitrarily large size, but still finite) to having an *infinite* union/product? Can we make necessary conclusions? Can we find counterexamples?

The main idea is that "passing to a limit" does create *some* mathematical object, but we can't necessarily pre-suppose that this object has the *exact same properties* as all of the objects in the sequence that defines that object.

Think about the finite sets $[n]$, for every n. Each of them is finite, but "in the limit" we get $\mathbb{N}$ which is *not* finite. So yes, we do get some object (another set), but it doesn't have to have the same properties.

The important theorem above shows that passing to the limit in the *union* definitely preserves countability. As we will see below in the next section, the *product* definitely does **not** preserve countability. (In fact, even an infinite product of finite sets is uncountable. Yikes!)

A similar notion appears in calculus. We promised we would not use calculus, but there is such a natural relationship between these ideas, so we feel compelled to mention an easy example. If you don't get anything out of this, no worries; if you do, though, try to remember this connection and think about how it might fundamentally change your view of everything you learned in calculus.)

Consider a *limit*, something like

$$\lim_{x \to \infty} \frac{1}{x} = 0$$

In what sense is this limit **equal** to 0? Why would we, as mathematicians over the years, choose to *define* limits in this way? Formally, this limit makes sense because of the quantified definition of a limit. Let P be the set of positive real numbers. Then the definition of limit (applied to this example) says

$$\forall \varepsilon \in P \mathbf{.}\ \exists M \in \mathbb{N} \mathbf{.}\ \forall n \in \mathbb{N} \mathbf{.}\ \left(n > M \implies \left|\frac{1}{x}\right| < \varepsilon\right)$$

That is, for any small positive threshold ($\varepsilon > 0$), we can find a specific cutoff point (a large natural number M that depends on ε somehow) such that, for every point *after* M, the function $\frac{1}{x}$ falls within that ε-threshold of the limit point, zero.

Notice that this is *very different* than saying some nonsense like "$\frac{1}{\infty} = 0$" That's **not** what's going on. We never actually get to "plug in" the end of the limit and evaluate it. The limit is defined in terms of quantifications, some things that are happening for *arbitrarily large* values, but not for an *infinite* value.

7.6.4 Uncountable Sets

To start our discussion of uncountable sets, let's prove a result we've mentioned already. Specifically, we will prove that a countably infinite *Cartesian product* of sets is uncountbly infinite. Notice that we don't even need to have the sets be infinite: we can make them all finite with size 2! We will use this result in the next part to demonstrate some examples of uncountable sets, including a familiar set we already know ...

An Uncountable Cartesian Product

Theorem 7.6.26. *A countably infinite Cartesian product of sets with just two elements is uncountably infinite. That is,*

$$\{0,1\}^{\mathbb{N}} = \{0,1\} \times \{0,1\} \times \{0,1\} \times \cdots$$

is uncountably infinite.

Proof. AFSOC that this set $\{0,1\}^{\mathbb{N}}$ is actually countably infinite. This means we can find a bijection between this set and $\mathbb{N}$; that is, we can identify a correspondence between all the elements of this set and all of the natural numbers. Thus, there is a *1st* element of this set, the element that corresponds to 1; there is a *2nd* element of this set, the element that corresponds to 2; and so on.

We don't know exactly what these elements are, we are just guaranteed that this correspondence exists. Still, we can write out all the elements y_i of $\{0,1\}^{\mathbb{N}}$ in a list. Each y_i is an ordered, infinite list of 0s and 1s, so we can write them like this:

$$\begin{array}{rcl}
1 & \leftrightarrow & (a_{1,1}\ ,\ a_{1,2}\ ,\ a_{1,3}\ ,\ a_{1,4}\ ,\ a_{1,5}\ ,\ \ldots) = y_1 \\
2 & \leftrightarrow & (a_{2,1}\ ,\ a_{2,2}\ ,\ a_{2,3}\ ,\ a_{2,4}\ ,\ a_{2,5}\ ,\ \ldots) = y_2 \\
3 & \leftrightarrow & (a_{3,1}\ ,\ a_{3,2}\ ,\ a_{3,3}\ ,\ a_{3,4}\ ,\ a_{3,5}\ ,\ \ldots) = y_3 \\
4 & \leftrightarrow & (a_{4,1}\ ,\ a_{4,2}\ ,\ a_{4,3}\ ,\ a_{4,4}\ ,\ a_{4,5}\ ,\ \ldots) = y_4 \\
5 & \leftrightarrow & (a_{5,1}\ ,\ a_{5,2}\ ,\ a_{5,3}\ ,\ a_{5,4}\ ,\ a_{5,5}\ ,\ \ldots) = y_5 \\
 & \vdots &
\end{array}$$

Every value $a_{i,j}$ is either 0 or 1. The i tells us which natural number we correspond to (i.e. the *vertical* position in the list) and the j tells us which coordinate we are in (i.e. the *horizontal* position in the list).

Since we have assumed the correspondence is a bijection, we know that this list contains *all* of the elements of $\{0,1\}^{\mathbb{N}}$. To complete the contradiction argument, we will construct an element of $\{0,1\}^{\mathbb{N}}$ that is guaranteed to **not** appear in this list! (This is a version of Cantor's Diagonalization Argument.)

Let's define the object $x = (x_1, x_2, x_3, \ldots)$ by saying

$$x_i = \begin{cases} 0 & \text{if } a_{i,i} = 1 \\ 1 & \text{if } a_{i,i} = 0 \end{cases}$$

That is, we are constructing x by going down the *main diagonal* of the grid of elements (so we see all of the elements $a_{i,i}$) and *switching the value* from a 1 to a 0, or vice-versa.

The following diagram is a *specific example* of how to do this, and is not part of this more general proof. However, we are including it for the sake of illustration:

$$\begin{aligned}
1 &\leftrightarrow \big(\textcircled{1}\,,\ 1\,,\ 0\,,\ 0\,,\ 1\,,\ \ldots\big) = y_1 \\
2 &\leftrightarrow \big(1\,,\ \textcircled{0}\,,\ 0\,,\ 0\,,\ 1\,,\ \ldots\big) = y_2 \\
3 &\leftrightarrow \big(0\,,\ 0\,,\ \textcircled{1}\,,\ 1\,,\ 0\,,\ \ldots\big) = y_3 \\
4 &\leftrightarrow \big(1\,,\ 1\,,\ 0\,,\ \textcircled{1}\,,\ 1\,,\ \ldots\big) = y_4 \\
5 &\leftrightarrow \big(0\,,\ 1\,,\ 1\,,\ 1\,,\ \textcircled{0}\,,\ \ldots\big) = y_5 \\
&\ \ \vdots \\
x &= \Big(\textcircled{0}, \textcircled{1}, \textcircled{0}, \textcircled{0}, \textcircled{1}, \ldots\Big)
\end{aligned}$$

Why would we choose to do this? Well, think about whether or not the object x could possibly belong to the list of elements above.

- Is $x = y_1$? No, because x is different from y_1 in their first coordinates. (In our example $x_1 = 0$ because $y_{1,1} = 1$.)
- Is $x = y_2$? No, because x is different from y_2 in their second coordinates. (In our example, $x_2 = 1$ because $y_{2,2} = 0$.)
- Is $x = y_3$? No, because x is different from y_3 in their third coordinates. (In our example, $x_3 = 0$ because $y_{3,3} = 1$.)

In general, for an arbitrary $i \in \mathbb{N}$, we can guarantee that x and y_i differ in the i-th coordinate. Accordingly, **none** of the y_i objects can be equal to this new object x. That is,

$$\big(\forall i \in \mathbb{N}\text{.}\ x_i \neq y_{i,i}\big) \implies \big(\forall i \in \mathbb{N}\text{.}\ x \neq y_i\big)$$

But the way we defined x, it is just an ordered, infinite list of 0s and 1s, so it is definitely an element of $\{0,1\}^{\mathbb{N}}$, itself.

This is a contradiction. We assumed we could list all the elements of our set, but we then used this ordering to construct an element of our set that definitely does not appear in the list. ⨳

Therefore, $\{0,1\}^{\mathbb{N}}$ is uncountably infinite. □

Note: This is a very slick argument. It's one of my favorite proofs in all of mathematics. Cantor was a genius for coming up with it and, what's even more interesting, it's actually fairly simple and memorable, as well. We belive taht you won't forget this "go down the main diagonal and switch the values" argument. The fact that we could even sumamrize the whole proof in nine words

like that is further indication of its brilliance.

Corollary: A countably infinite product of any sets with at least two elements each is uncountably infinite.

(Note: We really only need to say that none of the sets in the product are empty and that only finitely many of them are allowed to have exactly one element.)

Examples

You might be wondering now: what types of sets are uncountably infinite? Do we know any? Sure we do! Here are some examples.

Example 7.6.27. **The set of all infinite binary strings:**

You may have noticed that the set we used in the proof above—namely $\{0,1\}^{\mathbb{N}}$—is "essentially" the set S of infinite binary strings! An element of $\{0,1\}^{\mathbb{N}}$ is an infinitely-long ordered list of coordinates, each of which is 0 or 1. An element of S is an infinitely-long ordered list of 0s and 1s, but just without the parentheses and commas. As such, there is a very natural bijection between the two (just drop the parentheses and commas, or throw them back in), so we will identify these two sets as the same.

We saw above in Example 7.6.25 that the set of all finite binary strings is countably infinite. This latest result shows that the set of all infinite binary strings is uncountably infinite. An alternate proof of this fact involves finding a bijection between S and $\mathcal{P}(\mathbb{N})$, and then applying Cantor's Theorem that says $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})|$. (See Exercise 7.8.33 for these details.)

Example 7.6.28. $\mathbb{R}$ **is uncountably infinite:**

This is our first example of a standard set of numbers that is uncountably infinite. We can use the above result to prove this fact.

This claim makes some intuitive sense, since it "looks like" the real number line is "so much bigger" than just $\mathbb{N}$ or $\mathbb{Z}$. But we also saw that $\mathbb{Q}$ is countably infinite, and there are tons of rational numbers scattered across the real number line; in fact, between *any two real numbers* there lies infinitely many rationals!

What we will see now is that, yes, it is true that $\mathbb{R}$ is uncountably infinite. Furthermore, we will even show that $\mathbb{R}$ and $\mathcal{P}(\mathbb{N})$ are of the same "size" of infinity; that is, we will show $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$. (Remember that this is way more informative than just saying both sets are uncountable; there are many levels of uncountably infinite sets, we are just choosing not to talk about them too much so we don't hurt our brains.)

Morally speaking, the idea behind showing $\mathbb{R}$ is uncountably infinite, first of all, is to relate $\mathbb{R}$ to the set $\{0,1,2,3,4,5,6,7,8,9\}^{\mathbb{N}}$. Every real number can be expressed in decimal notation, which is just some ordered list of countably infinite many digits. There's a decimal point in there somewhere, and there are

issues like $0.999999\cdots = 1$, but those aren't huge deals. Since we already saw that even a "small" set like $\{0, 1\}$ yields an uncountable set when we take its product infinitely many times, then certainly a "bigger" set, like $\{0, 1, \ldots, 9\}$ will also give an uncountable set, even factoring in those issues. This is the intuitive argument you can carry around in your head and use to explain the result to your friends. (In fact, this is the argument you will find in most textbooks, as well.)

More formally, we can just prove that $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$. This stronger result implies that $\mathbb{R}$ is uncountably infinite (because Cantor's Theorem tells us $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})|$.) To do this, we will consider the set

$$I = \{y \in \mathbb{R} \mid 0 \leq y \leq 1\}$$

which is the interval $[0, 1] \subseteq \mathbb{R}$. We will show that

$$|\{0, 1\}^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})| = |I|$$

and then apply some results about bijections between intervals and $\mathbb{R}$.

Consider the function $f_1 : \{0, 1\}^{\mathbb{N}} \to I$ that takes in an infinite binary string, puts a decimal point in front of all the 0s and 1s, and says, "Evaluate this number as a **decimal** expansion".

As an example, consider the element that is $(1, 1, 0, 0, 1, 0, \ldots)$ where the rest are 0s. Then

$$f_1(1, 1, 0, 0, 1, 0, \ldots) = 0.110010\ldots_{\text{DEC}} = \frac{1}{10^1} + \frac{1}{10^2} + \frac{1}{10^5} = \frac{11001}{100000}$$

Notice that this is a function because any output is definitely a real number (since it has a decimal expansion; we just provided it) and it is somewhere between 0 and 1, since we put the decimal point in front. Furthermore, notice that f_1 is an **injection**; two different infinite binary strings must be different in some coordinate, so they yield two decimal expansions that differ somewhere and, thus, cannot be the same real number. This shows that $|\{0, 1\}^{\mathbb{N}}| \leq |I|$.

Consider the function $f_2 : \{0, 1\}^{\mathbb{N}} \to I$ that takes in an infinite binary string, puts a decimal point in front of all the 0s and 1s, and says, "Evaluate this number as a **binary** expansion".

As an example, consider the same element as above. Then

$$f_2(1, 1, 0, 0, 1, 0, \ldots) = 0.110010\ldots_{\text{BIN}} = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^5} = \frac{25}{32}$$

Notice that this is a function because any output is definitely a real number; just evaluate the resulting sum of fractions and it yields a real number between 0 and 1 (and even if the series is infinite, it is guaranteed to converge). For example, the input of all 0s yields 0 as an output, and the input of all 1s yields 1 as an output since

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \sum_{k \in \mathbb{N}} \frac{1}{2^k} = 1$$

Furthermore, notice that f_2 is a **surjection**. This fact hinges on some external knowledge about rational/irrational numbers; specifically, it is true that any irrational number can be approximated by a sequence of *dyadic* rational numbers (rationals whose denominators are powers of 2). We won't state or prove these results, but we think that by playing around with some examples, you'll start to see why this works. In fact, do some Googling for binary expansions of irrational numbers and you'll find some interesting results.

Since f_2 is a surjection, this shows $|\{0,1\}^{\mathbb{N}}| \geq |I|$. Accordingly, we conclude that $|\{0,1\}^{\mathbb{N}}| = I$. We also know $|\mathcal{P}(\mathbb{N})| = |\{0,1\}^{\mathbb{N}}|$ (see Exercise 7.8.33), so we now know that $|I| = |\mathcal{P}(\mathbb{N})|$.

The last step is to prove that $|I| = |\mathbb{R}|$. Look at Exercise 5 in Section 7.5.4. There, you found a bijection between the set $J = \{y \in \mathbb{R} \mid -1 < y < 1\}$ and $\mathbb{R}$. It is easy to find a bijection between J and the set $K = \{y \in \mathbb{R} \mid 0 < y < 1\}$ (try it now!). This shows that $|\mathbb{R}| = |J| = |K|$. Furthermore, $K \subseteq I$ and they differ by only two elements, 0 and 1, so $|K| = |I|$. Finally, this shows that $|I| = |\mathbb{R}|$, so we conclude that

$$|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$$

Look at the two arguments we mentioned:

- Considering the set $\{0,1,2,3,4,5,6,7,8,9\}^{\mathbb{N}}$, and
- Considering the set $\{0,1\}^{\mathbb{N}}$

Both arguments involved some knowledge about decimal expansions (and binary expansions). It seems there is no easy way around this, so we hope that the results above are still convincing. In particular, you might want to play around with the idea that f_2 in the discussion above is a **surjection** but not an **injection**. Can you convince yourself of these claims? Can you convince someone else?

Theorems

Let's see one results about uncountable sets. Then, we will state a final theorem about infinite sets, in general, before moving onwards!

Lemma 7.6.29. *Suppose A is uncountably infinite and B is countably infinite, and $B \subseteq A$. Then $A - B$ is uncountably infinite.*

(Note: We don't *need* to assume that $B \subseteq A$ here. If this were not the case, we would just consider A and $B \cap A$ as the sets, instead.)

Proof. Left for the reader as Exercise 5 in Section 7.6.5.

(**Hint:** Use a *contradiction* argument ...) □

Characterizing a Set as Infinite

To define **infinite** sets, we first defined **finite** sets, and then declared any set to be infinite if it is *not* finite. The following thereom shows us that we could have defined **infinite** in a different way. Namely, we can say a set is infinite if and only if we can find a bijection to a proper subset of itself. First, let's state and prove this helpful lemma; we will need it in the proof of the theorem below.

Lemma 7.6.30. *Let A be any set. Then, A is infinite $\iff$ there exists $B \subset A$ such that B is countably infinite.*

Proof. The $\Longleftarrow$ direction is obvious. If A is bigger than some infinite set, it is also infinite.

The $\Longrightarrow$ direction is more interesting. Suppose A is infinite. Let $\star \in A$ be some special element. We will take it out of consideration and construct a set B that is countably infinite and does not contain $\star$ as an element. This will guarantee $B \subset A$, with $B \neq A$.

Consider $A_1 = A - \{\star\}$. This set is also infinite, so we can choose some element $b_1 \in A_1$.

Consider $A_2 = A_1 - \{b_1\} = A - \{\star, b_1\}$. This set is also infinite, so we can choose some element $b_2 \in A_2$.

Consider $A_3 = A_2 - \{b_2\} = A - \{\star, b_1, b_2\}$. This set is also infinite, so we can choose some element $b_3 \in A_3$.

We can continue this process forever. Define $B = \{b_1, b_2, b_3, \dots\}$. (Note: we *are* "passing to a limit" here, but this is acceptable because we are not using this to "preserve" any properties of B. We are merely *constructing* the object B.)

Notice that B is countably infinite because there is an obvious bijection with $\mathbb{N}$. □

With this lemma in hand, we can state and prove the next result:

Theorem 7.6.31. *Let A be any set. Then, A is infinite $\iff$ there exists $B \subset A$ such that there exists $f : A \to B$ that is bijective.*

Proof. ($\Longrightarrow$) Suppose A is infinite. We must identify a proper subset $B \subset A$ and a bijection $f : A \to B$.

Since $A \neq \varnothing$, take any $x \in A$. Consider $B = A - \{x\}$. Notice $B \subset A$.

We want to show there is a bijection $f : A \to B$.

By Lemma 7.6.30 above, we know we can find a countably infinite strict subset $C \subset B$. (Note: A is infinite, so $B = A - \{x\}$ is also infinite, since we only removed one element. If you need more convincing, AFSOC B is finite, so it has some size; what, then, is the size of A?)

Since C is countably infinite, we can list the elements of C as $\{y_1, y_2, y_3, \dots\}$.

(Note: The idea is that there exists some bijection $g : \mathbb{N} \to C$, so we can let $y_1 = g(1)$ and $y_2 = g(2)$ and so on.)

Define $f : A \to B$ by

$$\forall y \in A. \quad f(y) = \begin{cases} y & \text{if } y \neq y_i \text{ for all } i \in \mathbb{N} \text{ and } y \neq x \\ y_1 & \text{if } y = x \\ y_{i+1} & \text{if } y = y_i \text{ for some } i \in \mathbb{N} \end{cases}$$

This is a bijection because we can identify its inverse function $F : B \to A$, which is

$$\forall z \in B. \quad F(z) = \begin{cases} z & \text{if } z \neq y_i \text{ for every } i \in \mathbb{N} \\ x & \text{if } z = y_1 \\ y_{i-1} & \text{if } z = y_i \text{ for some } i \in \mathbb{N} - \{1\} \end{cases}$$

We will leave it as an exercise for the reader to verify that $F = f^{-1}$. (Draw a picture to intuitively convince yourself, at the very least.)

($\Longleftarrow$) This direction claims that infinite sets are the *only* sets that have this property. We will prove this claim by contrapositive. That is, we will show that any finite set *cannot* have a bijection to a proper subset.

Suppose A is finite. This means it has a (unique) size, say $n \in \mathbb{N}$. Consider an arbitrary proper subset $B \subset A$. WWTS there cannot exist a bijection from A to B.

AFSOC there is such a bijection $f : A \to B$. Since B is finite and $B \subset A$, B has some size $m < n$. Thus, there is a bijection $g : B \to [m]$. Composing these bijections, we get a bijection $h : A \to [m]$. Thus, $|A| = n$ and $|A| = m$, so $m = n$. However, we also know $m < n$. This is a contradiction. ⨳

(Note: we can also make this argument via the *Pigeonhole Principle*, which we haven't yet discussed but will soon. Essentially, we can't have a bijection $p : [n] \to [m]$ when $n > m$ because there are "too few boxes" in which to stuff n "pigeons".) □

In the context of solving a problem, perhaps you'll want to argue that some set is infinite. Rather than proving that you cannot possibly find a bijection to *any* finite set, considering using this theorem! If you can identify a proper subset and a bijection, then you have accomplished your goal, with the help of this result.

7.6.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) When is a set **finite**?

(2) What are two ways to characterize when a set is **infinite**?

(3) What is the difference between **countably** and **uncountably** infinite? Give two examples of each type.

(4) Given two countably infinite sets, A and B, what set operations can we perform on them that are *guaranteed* to yield a countably infinite set? Might any set operations on them yield a *finite* set?

(5) Is $\mathbb{R} \times \mathbb{N}$ countably or uncountably infinite? What about $\mathbb{R} - \mathbb{N}$?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Prove Proposition 7.6.9. That is, prove: If A and B are finite sets, then

$$|A \cup B| = |A| + |B| - |A \cap B|$$

(2) Prove Corollary 7.6.10. That is, prove:

If $A_1, \ldots, A_n$ are finite and pairwise-disjoint, then

$$|A_1 \cup \cdots \cup A_n| = |A_1| + \cdots + |A_n|$$

(3) Find the flaw in the following "spoof" that $\mathbb{R}$ is countably infinite:

> Let $S \subseteq \mathbb{R}$ be the set defined by $S = \{y \in \mathbb{R} \mid 0 \leq y < 1\}$.
>
> For every $x \in S$, define the set $A_x = \{x + z \mid z \in \mathbb{Z}\}$.
>
> (For example, $A_{1/2} = \{\ldots, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\}$.)
>
> Since $\mathbb{Z}$ is countably infinite, each set A_x is countably infinite.
>
> Also, notice that
>
> $$\mathbb{R} = \bigcup_{x \in S} A_x$$
>
> This is a union of countably infinite sets, so $\mathbb{R}$ is also countably infinite.

Be sure to point out any particular step that is incorrect, as well as *why* that step is incorrect. Ideally, you should point out why the ultimate conclusion of the spoof is incorrect, but without just explicitly stating "$\mathbb{R}$ is uncountable because we proved that". *Why* is the incorrect step a misuse of a result, *and* why is the conclusion of that particular step invalid?

(4) For each of the following desired situations, provide an example or state that it is impossible.

For example, if the situation were "Finite sets A and B such that $A \cup B$ has size 4", an answer might be "Consider $A = \{1, 2\}$ and $B = \{3, 4\}$." If the situation were, "For every $x \in \mathbb{N}$, an infinite set S_x, such that $\bigcup_{x \in \mathbb{N}} S_x$ is finite", the answer would be "Impossible".

There is no need to *prove* your answers here; a good example should suffice.

(a) An uncountably infinite set A and a countably infinite set B such that $A \cap B$ is finite.

(b) Uncountably infinite sets C and D such that $C - D$ is countably infinite.

(c) Uncountably infinite sets E and F such that $E - F$ is uncountably infinite.

(d) For every $x \in \mathbb{N}$, a countably infinite set S_x, such that $\bigcup_{x \in \mathbb{N}} S_x$ is uncountably infinite.

(e) For every $y \in \mathbb{R}$, a countably infinite set T_y, such that $\bigcup_{y \in \mathbb{R}} T_y$ is countably infinite.

(5) Prove Lemma 7.6.29. That is, suppose A is uncountably infinite and $B \subseteq A$ is countably infinite; prove that $A - B$ is uncountably infinite.

Use this result to explain why the set of *irrational* real numbers is uncountably infinite.

7.7 Summary

Now, we have fully explored **functions** and their related properties! We saw that a function is just a relation with a particular property. This desired property corresponds to how we usually think of a function as having an "output" for every possible "input". We formalized these notions mathematically by defining terminology like domain, codomain, and image. Further properties of functions include injectivity and surjectivity. We saw many examples and non-examples of functions with these properties, and discussed how to prove/disprove these properties, relating back to our logical proof techniques.

The notion of a *bijection* has been particularly helpful and powerful. We related this to the notion of an *inverse* function. Specifically, we saw and proved that a function is bijective *if and only if* it has an inverse! This made for an important result later on when we discussed **cardinality**, where "the bijection is king". The notion of "pairing off elements" helped us make sense of some of the more wild and counter-intuitive results about the "sizes of sets".

We characterized infinite sets as either countably infinite or uncountably infinite. However, we also proved the historically significant result that is Cantor's

Theorem, which shows that there are, in fact, infinitely-many *cardinalities*! For our purposes here, it was sufficient to distinguish these two types of infinite sets. We saw several examples of each, and proved some theorems about how to create sets of specific cardinalities from others. Ultimately, we find these results intriguing and mathematically instructive. From now on, though, we will be focusing on **finite** sets only.

7.8 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them. Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ★.

Problem 7.8.1. For each of the following "rules" and proposed domains and codomains, determine whether the "rule" defines a **well-defined function**. Explain your answer using examples, if necessary.

(a) Let $a : \mathbb{Z} - \{1\} \to \mathbb{R}$ be defined by $a(x) = \dfrac{x^2}{x-1}$.

(b) Let $b : \mathbb{Q} \to \mathbb{Q}$ be defined by $b(x) = \sqrt{|x|}$.

(c) Let $c : \mathbb{Z} \to \mathbb{Z}$ be defined on every input $x \in \mathbb{Z}$ by outputting an $s \in \mathbb{Z}$ such that $x \equiv s \bmod 3$.

(d) Let $d : \mathbb{N} \to \mathbb{N}$ be defined by $d(x) = \left\lfloor \dfrac{x}{10} \right\rfloor$.

(e) Let $e : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{Z})$ be defined by taking in a set of natural numbers and outputting the set of all integer multiples of the least element of that set.

Problem 7.8.2. Consider the sets $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$ and $\mathbb{R}^2 = \{(a, b) \mid a, b \in \mathbb{R}\}$.

Consider the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (xz, yz)$. Is f injective? Surjective? Prove your claims.

Problem 7.8.3. Define $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by $f(x) = x + 1$ and $g(x) = x^2 + x$.

Find formulas for the compositions $f \circ g$ and $g \circ f$. (Notice that they are *different*.)

Prove that *neither* of those compositions is injective.

Problem 7.8.4. Let $f : \mathbb{Z} \to \mathbb{Z}$ be given by $f(x) = 2x - 3$. Let $g : \mathbb{Z} \to \mathbb{N}$ be given by $g(z) = |z| + 4$.

What is the domain of $g \circ f$? What is the codomain?

Write down a rule that defines $g \circ f$. Is this function injective? Surjective?

What is $\text{Im}_{g \circ f}(\mathbb{Z})$?

Prove your claims.

Problem 7.8.5. Each of the following rules defines a function from $\mathbb{N} \times \mathbb{N} \to \mathbb{Z}$. For each, determine whether the resulting function is injective or surjective, or both, or neither. Prove your claims.

(a) $f_1(a, b) = a - b$

(b) $f_2(a, b) = 2a + 3b$

(c) $f_3(a, b) = a$

(d) $f_4(a, b) = a^2 - b^2$

(e) $f_5(a, b) = 2^a \cdot 3^b$

Problem 7.8.6. Define functions f_1, f_2, f_3, f_4 with domain $\mathbb{N}$ and codomain $\mathcal{P}(\mathbb{N})$ with the following properties, or else explain why the desired properties are **not** possible to achieve.

- f_1 is injective and not surjective
- f_2 is neither injective nor surjective
- f_3 is surjective and not injective
- f_4 is bijective

Problem 7.8.7. Consider the function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined by

$$\forall (x, y) \in \mathbb{Z} \times \mathbb{Z}. \quad f(x, y) = (y + 1, 3 - x)$$

Find a function F that is the inverse of f, and prove that it is. What does this tell you about the function f?

Problem 7.8.8. Define the set $S = \{x \in \mathbb{R} \mid 0 < x < 1\}$. Define the function $g : S \to \mathbb{R}$ by

$$g(x) = \frac{2x - 1}{2x(1 - x)}$$

Prove that $\text{Im}_g(S) = \mathbb{R}$.

(**Hint:** You'll need to use the Quadratic Formula.)

Problem 7.8.9. Suppose $f : A \to B$ and $g : B \to C$ are functions.

(a) Suppose f, g are surjections. Prove that $g \circ f : A \to C$ is also a surjection.

(b) Suppose f, g are injections. Prove that $g \circ f : A \to C$ is also an injection.

(c) Suppose f, g are bijections. Prove that $g \circ f : A \to C$ is also a bijection.

Problem 7.8.10. Suppose $f : A \to B$ and $g : B \to C$ are bijections. Define $h : A \to C$ to be $h = g \circ f$.

Prove that h is invertible and that $h^{-1} = f^{-1} \circ g^{-1}$.

(**Hint:** Use the Associativity of Function Composition.)

Problem 7.8.11. Let $f : A \to B$ and $g : B \to C$ be functions. Let $X \subseteq A$. Prove that $\text{Im}_{g \circ f}(X) = \text{Im}_g(\text{Im}_f(X))$.

Problem 7.8.12. Let $f : A \to B$ be a bijection, so $f^{-1} : B \to A$ is a function. Let $X \subseteq A$. Prove that $\text{Im}_f(X) = \text{PreIm}_{f^{-1}}(X)$.

Problem 7.8.13. Let A, B be sets and let $f : A \to B$ be a function. Suppose $X, Y \subseteq A$.

(a) Is it necessarily true that the following equality holds?

$$\text{Im}_f(X \cup Y) = \text{Im}_f(X) \cup \text{Im}_f(Y)$$

State your claim and prove it.

(b) Is it necessarily true that the following equality holds?

$$\text{Im}_f(X \cap Y) = \text{Im}_f(X) \cap \text{Im}_f(Y)$$

State your claim and prove it.

Problem 7.8.14. Let $f : A \to B$ be a function. Define the relation $\sim$ on B by saying, for any $x, y \in B$,

$$x \sim y \iff \text{PreIm}_f(\{x\}) = \text{PreIm}_f(\{y\})$$

Explain why $\sim$ is an equivalence relation.

What are the equivalence classes?

Supposing that f is surjective, what are the equivalence classes?

Problem 7.8.15. Let $f : A \to B$ be a function. Define the relation $\approx$ on A by saying, for any $x, y \in A$,

$$x \approx y \iff f(x) = f(y)$$

Is $\approx$ an equivalence relation? If so, prove it, and describe the equivalence classes. If not, provide a counterexample.

Now, suppose that f is an injection. Is $\approx$ an equivalence relation? If so, prove it, and describe the equivalence classes. If not, provide a counterexample.

Problem 7.8.16. Let $f : A \to B$ be a function, and let $X, Y \subseteq A$. Consider the claim that $\text{Im}_f(X) \cap \text{Im}_f(Y) \subseteq \text{Im}_f(X \cap Y)$. What is wrong with the following "spoof" of that claim?

> Let $z \in \text{Im}_f(X) \cap \text{Im}_f(Y)$. Since $z \in \text{Im}_f(X)$, this means $\exists a \in X$ such that $f(a) = z$. Since $z \in \text{Im}_f(Y)$, this means $\exists a \in Y$ such that $f(a) = z$. Since $a \in X$ and $a \in Y$, we this means $a \in X \cap Y$. Since $f(a) = z$, this means $z \in \text{Im}_f(X \cap Y)$.

Provide a counterexample to show that the claim is, in fact, False.

Problem 7.8.17. Prove/disprove whether $\mathcal{P}(\mathbb{N})$ and $\mathcal{P}(\mathbb{Z})$ have the same cardinality.

Problem 7.8.18. Fix an arbitrary $n \in \mathbb{N}$. Consider the set $[n] = \{1, 2, 3, \ldots, n\}$.

Let E be the set of subsets of $[n]$ that have an **even** number of elements (like $\varnothing$ or $\{1, 4\}$), and let O be the set of subsets of $[n]$ that have an **odd** number of elements (like $\{5\}$ or $\{1, 2, 3\}$).

Define a function $p : E \to O$ that is a **bijection**, and prove that it is a bijection.

(**Hint:** Write out some small cases, where $n = 1$ and $n = 2$ and $n = 3$. Then try to generalize.)

Problem 7.8.19. Prove Lemma 7.6.16.

Hint: Try using a similar idea to our proof of Theorem 7.6.7: use the size of B to "bump up" the bijection between A and $\mathbb{N}$ by a certain amount.

Problem 7.8.20. Prove Corollary 7.6.19. That is, suppose A and B are countably infinite sets; prove that $A \cup B$ is countably infinite by applying Lemma 7.6.18 to appropriately-chosen sets.

Problem 7.8.21. Look back at Example 7.6.13. There, we defined $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by setting

$$\forall (x, y) \in \mathbb{N} \times \mathbb{N}. \quad f(x, y) = 2^{x-1}(2y - 1)$$

Prove that f is **injective**.

Problem 7.8.22. Prove Corollary 7.6.21. That is, suppose we have finitely-many sets—$A_1, A_2, \ldots, A_n$—where each set is countably infinite; prove that

$$A_1 \cup A_2 \cup \cdots \cup A_n$$

and

$$A_1 \times A_2 \times \cdots \times A_n$$

are both also countably infinite.

Problem 7.8.23. Consider the set A, defined by

$$A = \{(a, b) \in \mathbb{N} \times \mathbb{N} \mid a \leq b\}$$

Prove that A is countably infinite in two ways:

(1) By writing A as a union of sets and citing a result.

(2) By finding an explicit bijection between A and a countable set of your choosing.

Problem 7.8.24. Define $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by setting

$$g : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \qquad \forall (x, y) \in \mathbb{N} \times \mathbb{N}. \quad g(x, y) = (x + y)^2 + x$$

Prove that g is (a) injective and (b) not surjective.

Problem 7.8.25. Let A, B, C be sets. Let $f : A \to B$ and $g : B \to C$ and $h : B \to C$ be functions.

(a) Suppose $g = h$. Is it necessarily True that $g \circ f = h \circ f$? Prove or disprove this claim.

(b) Suppose $g \circ f = h \circ f$. Is it necessarily True that $g = h$? Prove or disprove this claim.

Problem 7.8.26. Let A, B be finite sets, with $|A| = |B| = n$. Suppose $f : A \to B$ is a function. Prove that

$$f \text{ is injective} \iff f \text{ is surjective}$$

Problem 7.8.27. Consider the following claim:

> Suppose $f : A \to B$ and $g : B \to C$ are functions. Suppose $g \circ f : A \to C$ is injective.
>
> Then g is also injective.

What is wrong with the following "spoof" of this claim?

> Suppose $g \circ f$ is an injection. We want to show g is an injection.
>
> Let $x, y \in B$ be given. Suppose $g(x) = g(y)$.
>
> We know $\exists a, b \in A$ such that $f(a) = x$ and $f(b) = y$.
>
> Since g is a well-defined function, this means $g(f(a)) = g(x)$ and $g(f(b)) = g(y)$.
>
> Since $g \circ f$ is injective and $g(f(a)) = g(f(b))$, this means $a = b$.
>
> Since f is a well-defined function, then $f(a) = f(b)$.
>
> This means $x = y$. Thus, g is injective.

Also, find a counterexample that shows the claim's conclusion is incorrect.

Problem 7.8.28. Let $a, b \in \mathbb{R}$ be arbitrary and fixed. Suppose $a^2 + b^2 \neq 0$.

Consider the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ defined by

$$\forall (x, y) \in \mathbb{R} \times \mathbb{R}\text{.} \qquad f(x, y) = (ax - by, bx + ay)$$

Prove that f is a bijection by finding its inverse and proving that inverse is correct.

Problem 7.8.29. Let A and B be finite sets and suppose $|A| = |B|$.

Suppose $f : A \to B$ is a function that is injective.

Prove that f must also necessarily be surjective by showing $\text{im}_f(A) = B$.

Problem 7.8.30. Let $k \in \mathbb{N} - \{1\}$ be given. Define

$$S_1 = \{X \in \mathcal{P}(\,[k]\,) \mid k \notin X\}$$

and

$$S_2 = \{X \in \mathcal{P}(\,[k]\,) \mid k \in X\}$$

(a) Prove that the sets S_1 and S_2 form a partition of $\mathcal{P}(\,[k]\,)$.

(b) Define a function $f_1 : S_1 \to \mathcal{P}(\,[k-1]\,)$ that is a bijection and prove that it is.

(c) Define a function $f_2 : S_2 \to \mathcal{P}(\,[k-1]\,)$ that is a bijection and prove that it is.

(d) Use what you proved in (a) and (b) and (c) to write an **induction** proof that $\mathcal{P}(\,[n]\,)$ has 2^n elements, for every $n \in \mathbb{N}$.

Note: Because of the restriction $k \geq 2$ above, make $n = 1$ your base case, use $n = k \geq 1$ in your Induction Hypothesis, and prove the claim for $n = k + 1$ in the Induction Step.

Problem 7.8.31. Let A, B, C, D be sets, and suppose $A \cap B = C \cap D = \varnothing$. Suppose $f : A \to B$ and $g : C \to D$ are bijections.

Define the piece-wise function $h : A \cup B \to C \cup D$ by setting

$$\forall x \in A \cup B\text{.}\ h(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B) \end{cases}$$

Explain why h is a well-defined function. Then, prove it is a **bijection**.

Problem 7.8.32. In this problem, you will prove that whenever A and B are finite with $|A| = a$ and $|B| = b$, it follows that $|A \times B| = ab$. This will be structured as a "double induction" proof on the two variables $a, b \in \mathbb{N}$.

(a) Show that $\big|\,[1] \times [1]\,\big| = 1$. (This is very, very easy, but necessary.)

(b) Suppose $n \in \mathbb{N}$ and $\big|\,[1] \times [n]\,\big| = n$. Show that $\big|\,[1] \times [n+1]\,\big| = n+1$.

(c) Explain why (a) and (b) have shown that $\forall n \in \mathbb{N}\text{.}\ \big|\,[1] \times [n]\,\big| = n$.

(d) Suppose $k \in \mathbb{N}$ and suppose $\forall n \in \mathbb{N}\text{.}\ \big|\,[k] \times [n]\,\big| = kn$. Show that $\forall n \in \mathbb{N}\text{.}\ \big|\,[k+1] \times [n]\,\big| = (k+1)n$.

(e) Explain why (c) and (d) have shown that $\forall k, n \in \mathbb{N}\text{.}\ \big|\,[k] \times [n]\,\big| = kn$.

(f) Explain why (e) proves the result stated in the problem description above.

Problem 7.8.33. Let S be the set of all infinite binary strings. (That is, elements of S are infinitely-long strings of 0s and 1s.)

Find a bijection between S and $\mathcal{P}(\mathbb{N})$. Use this to prove that S is uncountably infinite.

Problem 7.8.34. For each of the following sets, you are given its cardinality. Prove that the given cardinality is correct by finding a bijection to a relevant set and/or citing a result.

(**Hint:** If you don't use some kind of inductive argument, your proof might not be rigorous enough . . .)

(a) A is the set of all functions from $\mathbb{N}$ to $\mathbb{N}$. Show that A is uncountably infinite.

(**Hint:** Compare A with the set S of all functions from $\mathbb{N}$ to $\{1, 2\}$. Can you explain why S is uncountably infinite? What does this say about A? . . .)

(b) B is the set of all functions from $\mathbb{N}$ to $\mathbb{N}$ with the additional property that

$$\forall x \in \mathbb{N}\text{.}\ f(x+1) = f(x) + 1$$

Show that B is countably infinite.

(c) C is the set of all functions from $\mathbb{N}$ to $\mathbb{N}$ with the additional properties that

$$\begin{aligned} \forall x \in \mathbb{N}\text{.}\ f(x+1) &= f(x) + 1 \\ f(1) &= 42 \end{aligned}$$

Show that C is finite, and has only one element.

Problem 7.8.35. Look back at Example 7.6.14 where we argued (informally) that $\mathbb{N} \times \mathbb{N}$ is countably infinite by depicting the set as a lattice of points and describing a countably infinite path that covers all its points.

Formalize this argument by defining a function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ (or vice-versa) that achieves the path we described (or a similar one) and proving it is a bijection.

Problem 7.8.36. Prove Corollary 7.6.23. That is prove, that a countably infinite union of finite sets that are pairwise-disjoint is countably infinite.

Problem 7.8.37. . Consider Theorem 7.6.22, which states that a countably infinite union of countably infinite sets is also countably infinite. In our proof, we only considered the case where the given sets were *pairwise-disjoint*. In this problem, you should prove the general case, where the sets are not necessarily pairwise-disjoint.

(**Hint:** Consider the functions we used in our proof. Can you adapt them to find a *surjection* from $\mathbb{N} \times \mathbb{N}$ to the union of the sets?)

Problem 7.8.38. Consider the set S of all infinite binary strings. We proved that S is uncountably infinite before.

Consider the set $T \subseteq S$ that is the set of all infinite binary strings with only *finitely* many 1s.

In this problem, you will prove that T is, in fact, *countably* infinite!

(a) Consider the set $\mathbb{N}^k$ of all ordered k-tuples of natural numbers. (Note: $\mathbb{N}^1 = \mathbb{N}$ and $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$.)

Provide an inductive argument that shows that $\mathbb{N}^k$ is countably infinite for every $k \in \mathbb{N}$.

(Hint: This should be a pretty short proof. You should appeal to a result proven in lecture about Cartesian products of countably infinite sets.)

(b) For every $k \in \mathbb{N}$, let $T_k \subseteq T$ be the set of all infinite binary strings with *exactly* k 1s.

Find a bijective (or, at least, injective) function from T_k to $\mathbb{N}^k$. Explain why your function is well-defined and is a bijection (or injection).

(c) Use (b) to deduce that T_k is countably infinite. (Careful: If you found only an injection, you should also explain why T_k is not finite.)

(d) Express T as a union of sets and deduce that T is countably infinite.

(Hint: You'll need to apply an important Theorem from lecture.)

(Side note: Think about the consequences of this result. With a simple bijection, you can deduce that the set of all infinite binary strings with only finitely many 0s is *also* countably infinite. This means that the reason S, the set of *all* infinite binary strings, is *uncountably* infinite is completely tied to the set of strings with both infinitely many 1s and 0s. That set alone is big enough to make S uncountable!

Problem 7.8.39. (a) Let $n \in \mathbb{N}$. Consider the set

$$S = \{ f : [n] \to [n] \mid f \text{ is a bijection } \}$$

Show that S is *closed under composition*; that is, prove that

$$\forall f, g \in S \,.\, f \circ g \in S$$

(**Hint:** Cite a problem from this section of chapter exercises.)

(b) Consider the set

$$T = \Big\{ f : \mathbb{N} \to \mathbb{N} \mid f \text{ is a bijection and } \{i \in \mathbb{N} \mid f(i) \neq i\} \text{ is finite} \Big\}$$

Show that T is also *closed under composition*.

(c) Show that T is *closed under inverses*; that is, prove that

$$\forall f \in T \,.\; f^{-1} \text{ exists } \wedge\; f^{-1} \in T$$

(d) Consider the set

$$U = \big\{ f : \mathbb{N} \to \mathbb{N} \mid f \text{ is a bijection} \big\}$$

Prove that U is closed under inverses.

(e) Prove that

$$\forall f \in T \,.\; \forall g \in U - T \,.\; \big(f \circ g \notin T \;\wedge\; g \circ f \notin T \big)$$

(f) Find a counterexample to show that $U - T$ is **not** closed under composition.

(g) Furthermore, given $A \subseteq \mathbb{N}$ with A **finite**, find functions $f, g \in U - T$ such that

$$\{i \in \mathbb{N} \mid (f \circ g)(i) \neq i\} = A$$

(h) What are the cardinalities of S, T, U? If your answer is "finite", also state the size. If your answer is "infinite", also state whether it is *countable* or *uncountable* and prove your claim by finding a bijection to an appropriate set or citing a relevant result.

7.9 Lookahead

In the next chapter, we will study **combinatorics**, the mathematical branch of "counting things". We saw in the section on cardinality that many results about finite sets seemed rather intuitive. When we study combinatorics, we will be *describing* the elements of a set by characterizing what properties they have, rathern than simply stating them all or listing them. This will actually make it quite interesting (and sometimes very difficult!) to determine just how many elements we have described. Combinatorics is the study of techniques to determine the number of elements of a set with certain properties. We will state and prove some fundamental principles of counting (appealing to results from this chapter, in fact) and use them to build more advanced techniques and solve some interesting problems.

Chapter 8

Combinatorics: Counting Stuff

8.1 Introduction

The field of **combinatorics** is one of the most active and exciting areas of interest in modern mathematics. It is also sometimes known as "discrete math" to distinguish it from **analysis**, which studies more "continuous" notions like the real number line and functions defined on that set. In this chapter we will explore some of the fundamental ideas in combinatorics and apply them to solve interesting problems. Essentially, we will be learning interesting and useful principles about how to *count* the number of elements in finite sets where those elements are described in some way but not enumerated for us.

8.1.1 Objectives

The following short sections in this introduction will show you how this chapter fits into the scheme of the book. They will describe how our previous work will be helpful, they will motivate why we would care to investigate the topics that appear in this chapter, and they will tell you our goals and what you should keep in mind while reading along to achieve those goals. Right now, we will summarize the main objectives of this chapter for you via a series of statements. These describe the skills and knowledge you should have gained by the conclusion of this chapter. The following sections will reiterate these ideas in more detail, but this will provide you with a brief list for future reference. When you finish working through this chapter, return to this list and see if you understand all of these objectives. Do you see why we outlined them here as being important? Can you define all the terminology we use? Can you apply the techniques we describe?

By the end of this chapter, you should be able to ...

- State the Rules of Sum and Product, and use and combine them to construct simple counting arguments.
- Categorize several standard counting objects, as well as state corresponding counting formulas and understand how to prove them.
- State the meaning of binomial coefficients and evaluate their numerical formulae, know how to use them in counting arguments, and understand how to derive those numerical formulae.
- Critique a proposed counting argument by properly demonstrating if it is an undercount or overcount.
- Prove combinatorial identities by constructing "counting in two ways" proofs.
- Understand various formulations of selection with repetition, and use them to solve problems.
- State the Pigeonhole Principle and use it in counting arguments.
- State the Principle of Inclusion/Exclusion and use it in counting arguments.

8.1.2 Segue from previous chapter

In Chapter 7, we left off talking about the cardinality of sets, both finite and infinite. While many of the results about infinite sets are interesting and mathematically rich, that particular area can lead to some mind-bending and confusing areas that are, alas, beyond the scope of our current studies. For now, we will focus on finite sets. In particular, we will explore how some results about the cardinality of finite sets can be used to solve problems about "counting" mathematical objects. That is, we will explore how we can answer questions of the form "How many objects are there with property X?" This branch of mathematics is known as *combinatorics*. You can think of it as "the study of combinations of objects". While investigating this branch of mathematics, we will develop some new notation and definitions, prove and use some results about finite sets, and describe and study some particular objects that live in the field of combinatorics and computer science. Importantly, we will learn an entirely new proof technique based on counting objects!

8.1.3 Motivation

Think about playing poker. If you're unfamiliar with the game, just think of it as a simple system where two players receive a hand of 5 random cards each and then they compare to see who wins. Hands are ranked according to the following list, from best to worst:

- Straight Flush (five in a row of one suit), e.g. $T\clubsuit\ J\clubsuit\ Q\clubsuit\ K\clubsuit\ A\clubsuit$
- Four of a Kind, e.g. $3\clubsuit\ 3\spadesuit\ 3\heartsuit\ 3\diamondsuit\ 7\heartsuit$
- Full House (three of a kind and a pair), e.g. $4\clubsuit\ 4\spadesuit\ 4\diamondsuit\ 6\clubsuit\ 6\heartsuit$
- Flush (five of one suit), e.g. $2\heartsuit\ 5\heartsuit\ 8\heartsuit\ Q\heartsuit\ K\heartsuit$
- Straight (five in a row, not all the same suit), e.g. $8\diamondsuit\ 9\clubsuit\ T\diamondsuit\ J\heartsuit\ Q\heartsuit$
- Three of a Kind, e.g. $K\spadesuit\ K\heartsuit\ K\diamondsuit\ Q\heartsuit\ 9\clubsuit$
- Two Pair, e.g. $A\spadesuit\ A\heartsuit\ J\spadesuit\ J\diamondsuit\ 2\clubsuit$
- One Pair, e.g. $8\heartsuit\ 8\diamondsuit\ 2\spadesuit\ 5\clubsuit\ K\heartsuit$
- High Card, e.g. $Q\spadesuit\ J\clubsuit\ 9\diamondsuit\ 7\diamondsuit\ 2\diamondsuit$

Is this a fair game? If you've played poker before, and especially if you've played a lot, you've not only learned to accept this ranking system but you've also learned how to exploit it and make decisions. In Five Card Draw, if you're dealt 22345, should you keep the pair or go for the straight? Which is more likely to happen? Which will pay off more handsomely?

By our question, "Is this a fair game?", what we're really wondering is *why* the ranking is the way it is! Is drawing a flush actually rarer than a straight? Does it make sense that a full house loses to four of a kind? Why? How can we *prove* these results? To answer these questions, we will rephrase the questions in terms of *counting* instead of probability. We will ask how *many* distinct five card hands are flushes, how *many* are straights, and so on. This will allow us to compare them directly. Do you see how this relates to our work in the previous chapter, too? We will really be identifying the *cardinality* of the set of all poker hands that are flushes, for instance, and comparing it to the cardinalities of other sets of hands.

8.1.4 Goals and Warnings for the Reader

We will need to develop some notation and definitions to begin formulating a method to count the elements of particular finite sets, but we want to emphasize that this is really what is going on, overall: *combinatorics* is about counting the number of elements in finite sets using particular methods (which we will develop in this chapter). More specifically, we want to study these counting techniques in an abstract sense so that we can apply them in an *efficient* manner. Perhaps we could answer those poker questions we posed above by looking at all possible five card hands and making a tally mark every time we see a flush, say, but surely this will take way too long! There must be a better way! Well, of course, there is, and we will develop it soon enough in this chapter's first section.

We want to emphasize that we will be developing a new style of proof in this chapter, as well. More so than previous problems and techniques we've studied, proofs in combinatorics depend greatly on clarity and specificity of language.

Some of your proofs to the exercises in these sections may consist entirely of English sentences, with almost no mathematical symbols! This will seem strange at first, and might even seem to contradict the ideas we have emphasized so far about precision, clarity, and mathematical rigor. This is definitely not the case, though; combinatorics has a rigorous foundation in finite set theory, and we will work hard to point out this relationship whenever it is relevant. This property of combinatorics will also require you to be extra careful about your proof-writing style, ensuring that your words are chosen appropriately to be unambiguous and clear. More so than ever, be sure to reread your proofs after writing them, pretending that you are someone else, to make sure that the points you want to make actually come across in your proofs.

One final introductory point can be made by the following quote that a friend of mine stated once when we were talking about how to teach combinatorics. I found that it nicely summarized the sometimes strange transition from the the proofs we have been doing so far (that might feel rather formal) to combinatorics proofs (that might feel rather informal, in comparison).

> Finite cardinality is boring. That's not inconsistent with the fact that combinatorics is hard.

You might not know what that means now, but if you look back at this quote after working through this chapter, you'll understand what he was getting at. What this means is that, in an abstract and theoretical sense, finite cardinality *is* boring; all the results are what you'd expect them to be—like $|A \cup B| = |A| + |B|$ when $A \cap B = \varnothing$—and the techniques are all the same—find a bijection to an appropriate set. *Infinite* cardinality is far stranger and surprising—$|A \cup B| = |A| + |B|$ can be False, even if $A \cap B = \varnothing$, and even further, the addition $|A| + |B|$ is hard to make sense of, mathematically!

How does combinatorics differ, then? Well, in all of our work with combinatorics, we are given a finite set; the difference is that its elements are only *described* to us in some way. We are not *presented* with the elements of a set directly and asked to count them. (That would be easy: "One, two, three, . . . ") We have to come up with relevant and helpful strategies to identify how *many* objects have a certain prescribed list of properties. *That* is where the difficulty of combinatorics comes in. When we say, "Consider the set of all 5-card hands, as drawn from a standard deck of cards", you can immediately grasp the idea of that set, but you certainly can't picture *all* its elements laid out before you, let alone begin to count them one-by-one. In this sense, combinatorics is hard; this is also why it is incredibly interesting and popular!

8.2 Basic Counting Principles

8.2.1 The Rule of Sum

Look back at Theorem 7.6.7 that we proved in the previous chapter. It says that when we take two finite sets that are disjoint (i.e. they share no elements),

the size of their union is the sum of their individual sizes. This makes intuitive sense for finite sets, and we proved the result mathematically using A bijection. This result forms the basis for the first, fundamentally useful principle of combinatorics. Notice that this grounds us firmly in the principles of set theory.

Partitions

We start by recalling Definition 3.6.9, which was introduced in our discussion of sets.

Definition 8.2.1. *Let A be a set. A* **partition** *of A is a collection of sets that are pairwise disjoint and whose union is A.*

That is, a partition is formed by an index set I and non-empty sets S_i (defined for every $i \in I$) that satisfy the following conditions:

(1) For every $i \in I$, $S_i \subseteq A$.

(2) For every $i, j \in I$ with $i \neq j$, we have $S_i \cap S_j = \varnothing$.

(3) $\bigcup_{i \in I} S_i = A$

Essentially, a partition is a way of breaking a set into smaller sets that do not overlap. Let's look at a couple of examples before moving on.

Example 8.2.2. Let A be the set of people in the room currently. Let $I = \{1, 2\}$, and let S_1 be the set of left-handed people and let S_2 be the set of right-handed people. Then $S = \{S_1, S_2\}$ is a partition of A. Notice the distinction between writing "$\{S_1, S_2\}$ partition A", which is correct, and "S_1, S_2 partition A", which is not correct. What does it mean to say S_1, S_2 in this context? We really mean that those two sets, taken *together* as a collection, form a partition of A. This is why we must remember to write the elements S between brackets.

To be rigorous, we should *prove* why S is a partition of A. To do this, we point out that $S_1 \cap S_2 = \varnothing$ because everyone here is either left- or right-handed but not both. (Let's presume there are no "outlying cases" here, like truly ambidextrous people or people with no hands. If any such people are present, include them in a set S_3 and include that in our partition set S.) We also point out that $S_1 \cup S_2 = A$ because everyone in the room must be either left- or right-handed, so there *cannot* exist an element $x \in A$ that satisfies $x \notin S_1$ and $x \notin S_2$. This shows why S is a partition.

What if we wanted to partition the set of people in this room by separating them based on the first letter of their first name? Try to define this partition using mathematical notation like the previous example.

Example 8.2.3. Now, let's see a non-finite partition. Consider the set $A = \mathbb{N}$ and the index set $I = \mathbb{N}$. For every $i \in \mathbb{N}$, define the set

$$S_i = \{2i - 1, 2i\}$$

Is the set $S = \{S_i \mid i \in \mathbb{N}\}$ a partition of $\mathbb{N}$? We think so; let's investigate why. We could start by writing out what the first few sets look like (indeed, this is usually a good first strategy: just write out the first few cases and see what happens):

$$\begin{aligned} S_1 &= \{1, 2\} \\ S_2 &= \{3, 4\} \\ S_3 &= \{5, 6\} \\ &\vdots \end{aligned}$$

and so on. This looks like a partition of $\mathbb{N}$ so far, doesn't it? Let's prove that it truly is!

First, let's show that the sets S_i are *pairwise-disjoint* (i.e., any two of the sets share no elements). We prove this by contradiction. AFSOC that $\exists i, j \in \mathbb{N}$ with $i \neq j$ such that $S_i \cap S_j \neq \varnothing$. This means that (at least) one element of S_i is also an element of S_j; we find there are four possible cases for this situation:

1. $2i - 1 = 2j - 1$
2. $2i - 1 = 2j$
3. $2i = 2j - 1$
4. $2i = 2j$

The first and fourth cases immediately imply that $i = j$, by some simple algebra, which contradicts our given condition that $i \neq j$. The second and third cases are contradictions themselves because they involve an odd natural number and an even natural number being equal. In any case, we find a contradiction. Therefore, $\forall i, j \in \mathbb{N}$ with $i \neq j$, it's the case that $S_i \cap S_j = \varnothing$.

Second, let's show that the union of all of the S_i sets is $\mathbb{N}$. That is, let's prove

$$\bigcup_{i \in \mathbb{N}} S_i = \mathbb{N}$$

Remember that the set on the left-hand side consists of all of the elements x such that $\exists i \in I$ that satisfies $x \in S_i$. (Think about why this makes sense, even though I is infinite. This just means the union contains all of the elements that belong to at least one of the sets S_i.) Notice that for every $i \in \mathbb{N}$, the elements $2i - 1, 2i \in S_i$ are both natural numbers. Thus,

$$\mathbb{N} \supseteq \bigcup_{i \in I} S_i$$

Next, we prove the reverse set containment. Let $n \in \mathbb{N}$. We have two cases to consider. (1) If n is even, then $\exists k \in \mathbb{N}$ such that $n = 2k$. Thus, $n \in S_k$. (2) If n is odd, then $\exists \ell \in \mathbb{N}$ such that $n = 2\ell - 1$. Thus, $n \in S_\ell$. In either case, we have shown that $n \in \bigcup_{i \in I} S_i$.

Therefore, S is a partition of $\mathbb{N}$. In particular, it is an infinite partition.

Now we have seen an example of a finite and infinite partition.

(**Challenge question:** Can you identify an infinite partition of $\mathbb{N}$ such that all of the component sets of the partition are *also* infinite?)

Statement

For the remainder of the chapter, we will only consider *finite* partitions of finite sets. In particular, the Rule of Sum only applies in this specific case.

Proposition 8.2.4. *Let A be a finite set, let $n \in \mathbb{N}$, and let $S = \{S_i \mid i \in [n]\}$ be a finite partition of A. The* **Rule of Sum** *states that*

$$|A| = \sum_{i \in [n]} |S_i|$$

The Rule of Sum tells us that the size of a set can be found by partitioning it into a finite number of smaller sets and summing their sizes. Notice that this is precisely Corollary 7.6.10 that we saw last chapter in our discussion of finite sets! There, we asked you to prove this claim by induction, in Exercise 2 in Section 7.6.5. With this result in hand, we'll move on to see some examples.

Examples

Example 8.2.5. At Unique Activity University, every student is required to participate in *exactly* one varsity sport each year. Playing more than one would be too much of a time commitment, and not playing at all would make them lazy, so everyone plays exactly one of the following non-traditional-but-still-sports sports: golf, cricket, badminton, and chess. The athletic department released the following statistics about the rosters for each sport this year:

- Golf: 12 players
- Cricket: 18 players
- Badminton: 23 players
- Chess: 33 players

How many students attend UAU?

Okay, this is an easy example because we made sure to stipulate that the sports offered by the university form a partition of the set of students. (Compare that to the sentence, "The set of sports offered by the university is a partition of the set of students." Both are correct.) Thus, we can find the cardinality of S, the set of all students, by adding;

$$|S| = 12 + 18 + 23 + 33 = 86$$

A small university, indeed, as well as a bizarre one. Don't go there.

More interesting examples of applying the Rule of Sum will appear when we combine it with other counting principles. For now, it's a simple idea that governs how to count sets that can be broken into disjoint parts. In general, the hardest part about using the Rule of Sum is deciding *which* partition to apply it to, and being creative about that.

The next counting principle is just as, if not more, helpful but a little more intricate to define and prove.

8.2.2 The Rule of Product

Motivation

We'll motivate this principle via an example.

Example 8.2.6. Let's say we have three people in the room. We also have three stickers bearing the numbers 1, 2, and 3 on them (with one distinct number on each sticker). How many ways can we place these stickers on the three people? For the sake of argument, let's say the people are named Andy, Brendan, and Carl, conveniently abbreviated as A, B, and C. To answer this question, we can simply write out all of the sticker assignments in an organized manner to make sure we don't miss any. Specifically, we'll rank them in increasing order by Andy's assignment, then Brendan's, then Carl's: we have $(A, B, C) =$

1. $(1, 2, 3)$
2. $(1, 3, 2)$
3. $(2, 1, 3)$
4. $(2, 3, 1)$
5. $(3, 1, 2)$
6. $(3, 2, 1)$

Thus, there are 6 total ways to assign the stickers.

What if we have four people–Andy, Brendan, Carl, and Dave–and four stickers? Can we list all of those assignments? Sure, why not?

$$\begin{array}{cccc}
(1,2,3,4) & (1,2,4,3) & (1,3,2,4) & (1,3,4,2) \\
(1,4,2,3) & (1,4,3,2) & (2,1,3,4) & (2,1,4,3) \\
(2,3,1,4) & (2,3,4,1) & (2,4,1,3) & (2,4,3,1) \\
(3,1,2,4) & (3,1,4,2) & (3,2,1,4) & (3,2,4,1) \\
(3,4,1,2) & (3,4,2,1) & (4,1,2,3) & (4,1,3,2) \\
(4,2,1,3) & (4,2,3,1) & (4,3,1,2) & (4,3,2,1)
\end{array}$$

Okay, so there are 24 total ways to assign the stickers. What about with five people? I don't know about you, but my arm is getting tired writing out all of these assignments. There must be a better way to do this! Yes! This is where

the Rule of Product comes in to save the day. (Side note: You might notice a pattern to our list above; can you infer how we made sure we actually listed *all* possibilities? Could you write a little computer program that would generate all the possibilities, for any number of elements? Try it!)

Statement

We will actually make two separate statements of the Rule of Product. The first is an intuitive statement of when and how it applies and what it claims. The second is a more rigorous, mathematical statement that is rooted in the kind of set-theoretic language that we have been using all along. We emphasize that both definitions should, ideally, be understood; however, truly understanding the first one is more important, and the second is presented mostly because it is the one that can and will be rigorously proven.

Proposition 8.2.7. *Consider a process that is completed in n distinct steps. Assume that the i-th step, for every $i \in [n]$, has exactly w_i different ways to be completed; moreover, assume that this number $w_i \in \mathbb{N}$ does* not *depend on the choices made in the previous steps. Also, assume that no two distinct choices at any step yield the same outcome. Then the Rule of Product states that the total number of outcomes, N, of this n-step process is*

$$N = \prod_{i \in [n]} w_i$$

Let's relate this statement back to the previous example with the people and stickers before moving on and stating the Rule of Product more rigorously.

Example 8.2.8. We can think of assigning the stickers to Andy, Brendan, and Carl as a three-step process. Let's line up the three gentlemen in alphabetical order, left to right, then move along the row. At each step, we will place a sticker on the gentleman in front of us by choosing one that hasn't been assigned yet. In the first step, we approach Andy and have 3 possible stickers to place on him. In the second step, we approach Brendan and have 2 possible stickers to place on him. Notice that this is true *no matter what sticker was chosen for Andy*. We don't actually care *which* sticker was chosen for Andy—be it 1, or 2, or 3—merely that the *number* of choices we have when we face Brendan is *always* 2. In the third step, we approach Carl and find that we have only 1 sticker option, regardless of the previous two choices.

The Rule of Product tells us that the number of ways to complete this process is the product of those numbers of options at each step: $3 \cdot 2 \cdot 1 = 6$. This agrees with our "exhaustive list" procedure. Hooray!

What about with 4 people? Using the same kind of logic, we can see that there are $4 \cdot 3 \cdot 2 \cdot 1 = 24$ possible ways to complete the sticker-assignment process. Again, this agrees with our previous procedure. Double hooray!

What about with 5 people? Well, $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. We figured out something we didn't know yet. Triple hooray! With 6 people? With 7 people?

With n people, where $n \in \mathbb{N}$? We can answer all of these questions very easily and precisely now, thanks to the Rule of Product. Infinite hooray!

Tree Diagrams

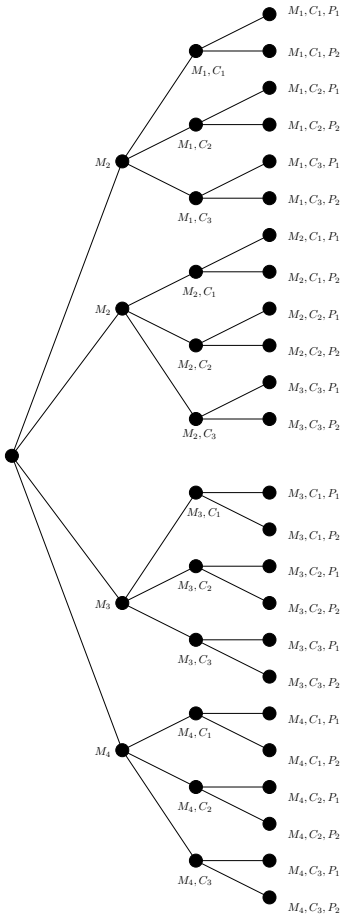
An interesting and helpful interpretation of the Rule of Product is evidenced by a *tree diagram*. The concept of a *tree* arises in the branch of mathematics known as *graph theory*, which studies mathematical objects consisting of vertices (dots) connected by edges (lines between the dots, where we only care about whether or not a line is *present*, and not on what it "looks like" when drawn on a piece of paper). A tree is a particular type of graph, and it arises commonly in computer science, as well, when studying *branching processes*. Within our context, we can use a tree to represent the decision points of a *procedure* whose end products will be counted by the Rule of Product. Furthermore, this method will provide some insight into the mathematically *rigorous* statement and proof of the Rule of Product. (We will leave these ultimate goals to the exercises, but for those of you who are interested and motivated to attempt them, we strongly encourage reading this section, as well; it will give you some intuition and guide you through those exercises.)

Example 8.2.9. Let's illustrate tree diagrams and how they relate to the Rule of Product via an example. Let's say we are planning our schedule for next semester. Based on our major and time constraints (and personal interests, of course), we must take exactly one class from each of three departments: mathematics, computer science, and philosophy. The number of courses available for us to take in each department does *not* depend on the selection we make in any other department; specifically, we have 4 mathematics courses to choose from, 3 computer science courses, and 2 philosophy courses, and any combination of courses will fit our schedule (provided each department is represented exactly once).

How might we apply the Rule of Product to our situation? We would need to define a *process* and the *steps* of that process, and then identify how many *choices* are available at each step. Naturally, the overall process here is identifying our course schedule for next semester. Since we are constrained to select (exactly) one course from each department, let us identify three steps: (1) select a mathematics course; (2) select a computer science course; (3) select a philosophy course. (Note: Does the *order* of these steps matter? What if we select a philosophy course first, instead? Will our process be fundamentally *different*? We think not, but make sure you see why before reading on.)

Next, let's represent the choices we can make at each step. Let's say the set of 4 mathematics courses available to take is $\mathcal{M} = \{M_1, M_2, M_3, M_4\}$, the set of 3 computer science courses is $\mathcal{C} = \{C_1, C_2, C_3\}$, and the set of 2 philosophy courses is $\mathcal{P} = \{P_1, P_2\}$. This immediately identifies for us the *number* of choices available at each step: (1) there are $|\mathcal{M}| = 4$ choices; (2) there are $|\mathcal{C}| = 3$ choices; (3) there are $|\mathcal{P}| = 2$ choices. Thus, the Rule of Product tells us there are 24 total course schedules we could create for next semester. But,

really, *why* is this true? What *are* those schedules? Let's represent them by a tree diagram!



Reading left to right in the diagram, we are following this three step procedure we established. The single vertex (or node) at the far left represents the start of our process—no decisions have been made—and the four edges (or branches) emerging from that vertex represent the four mathematics courses from which we can choose. We have labeled each edge with one of the elements of $\mathcal{M}$. No matter which one of those edges we follow (i.e. no matter which mathematics course we select), there are three edges emerging from the next vertex (i.e. we still have three computer science courses from which we can choose). We have labeled all of those edges with corresponding elements from $\mathcal{C}$. Following the same idea, every vertex in that column has two emerging edges which are labeled by corresponding elements from $\mathcal{P}$.

The benefit of this diagram is that we can see *exactly* what the 24 outcomes of this process are by following the labels on the edges. For instance, look at the vertex on the top of the far right column. This corresponds to selecting M_1 and

C_1 and P_1; alternatively, we can represent this as the ordered triple (M_1, C_1, P_1). Further down that column, we see a vertex corresponding to the ordered triple (M_2, C_3, P_1), for example. Every vertex has an ordered triple representation! What we are really doing when we apply the Rule of Product is identifying the cardinality of some *set* that is a *Cartesian product* of several constituent sets. The *process* corresponds to identifying elements of the constituent sets and arranging them in an ordered tuple. The Rule of Product tells us how many *ways* we can do this by identifying the *cardinality* of the product set consisting of *all* such tuples. In this specific example, we have

$$|\mathcal{M} \times \mathcal{C} \times \mathcal{P}| = |\mathcal{M}| \cdot |\mathcal{C}| \cdot |\mathcal{P}| = 4 \cdot 3 \cdot 2 = 24$$

Does this make more sense, now? Does this provide you any insight into how the Rule of Product actually *works*?

More Formal Statement

See Exercise 8.9.1, which asks for a proof of the following theorem. This is a more formal statement of what the Rule of Product is, mathematically. After the statement, we'll describe how it relates to the previous version.

Theorem 8.2.10. Rule of Product (Set-Theoretic Version)

Let $n \in \mathbb{N}$. Suppose that $\forall i \in \mathbb{N} \bullet T_i$ is a finite set. Then,

$$\left| \prod_{i \in [n]} T_i \right| = |T_1 \times T_2 \times \cdots \times T_n| = |T_1| \cdot |T_2| \cdots |T_n| = \prod_{i \in [n]} |T_i|$$

The relationship with the previously-stated rule of product is as follows. The elements of the set T_1 are the choices that can be made in Step 1 of the process. For every element of T_1, we define the set T_2 to be the set of choices that can be made in Step 2 of the process *after* that choice made in Step 1. By assumption, there is an equal *number* of such choices, regardless of the choice made in Step 1. Thus, it makes sense that the conclusion of the Theorem only incorporates $|T_2|$, since this value is well-defined. Likewise, T_3 is the set of choices for Step 3 that can follow the choices made in Steps 1 and 2, and by assumption, $|T_3|$ is well-defined.

In the end, we can describe an **outcome** of this process by an ordered n-tuple, where coordinate i is an element of the set T_i. Indeed, what that element could be does depend on what the previous coordinates are, but the **number** of choices for this element is independent of those prior choices. Since, in the end, we really only care about the **number** of possible outcomes, the result makes sense. Actually **listing** all of the outcomes would require a careful analysis of each step, seeing how a particular choice affects the choices in the next step (and the steps thereafter), but that's not the point of the result. This is why, essentially, the result amounts to proving that the size of a product of finite sets is equal to the product of their sizes.

Example: Applying the Rules of Sum and Product (Together)

Let's practice using these two combinatorics Rules. You'll also notice that we'll start abbreviating these rules as ROS and ROP, so that we can cite them easily. And yes, we do *need* to cite them when we use them!

Example 8.2.11. **License Plates:**

Suppose a license plate string consists of 6 or 7 positions, each of which is filled with a letter (from A to Z) or a digit (from 0 to 9).

(1) How many license plates are there?

We must partition based on the length of the string, whether it is 6 or 7.

Within each part, we have a 6 or 7 step process. At step i, we fill Position i in the string with one of the 36 options (there are 26 letters and 10 digits). By ROP, then, there are 36^6 strings of length 6 and 36^7 strings of length 7.

By ROS, then, there are $36^6 + 36^7$ total license plate strings.

(2) How many license plates have at most 1 digit?

We must partition based on whether there are 0 digits or 1 digit.

With 0 digits, each step in our process places a letter in the corresponding position. We either have 6 letters—yielding 26^6 possibilities—or 7 letters—yielding 26^7 possibilities, by ROP.

By ROS, there are 36^6 or 36^7 such outcomes.

With 1 digit, step 1a chooses which of the positions is filled with a digit, step 1b chooses the digit for that position, and the rest of the steps fill the remaining positions with letters only.

There are 6 choices for which position is a digit, then 10 choices for how to fill *that* position (wherever it is), and 26 choices each for the other positions. Applying ROS and ROP, we find there are $6 \cdot 10 \cdot 36^5$ or $7 \cdot 10 \cdot 36^6$ such outcomes.

In total, by ROS, there are

$$(36^6 + 6 \cdot 10 \cdot 36^5) + (36^7 + 7 \cdot 10 \cdot 36^6)$$

total outcomes.

(3) How many license plates have at least 2 digits?

We could follow the same method we used with the previous question, and partition this set of license plates into those with 3 digits, 4 digits, 5 digits, 6 digits, and 7 digits. We would then need to count each such set and add their sizes. But how many license plates have, say, 4 digits? With 6 positions to be filled, how many ways are there to choose 4 positions to be digits? This is where *binomial coefficients* will be helpful, soon enough

(after we have defined them and derived a formula).

Instead, let's take advantage of the work we just did! Let's partition the set of *all* license plates (call this set Y) into those with at most 1 digit (call this set X_1) and those with at least 2 digits (call this set X_2). Notice that this *is* a partition, so ROS tells us $|Y| = |X_1| + |X_2|$. Subtracting algebraically, this tells us the expression we want is

$$\begin{aligned} |X_2| &= |Y| - |X_1| \\ &= (36^6 + 36^7) - \left[(36^6 + 6 \cdot 10 \cdot 36^5) + (36^7 + 7 \cdot 10 \cdot 36^6)\right] \end{aligned}$$

by just substituting in the expressions we've already derived. How convenient!

In general, this is a good strategy: to count a set, we can count its *complement* (i.e. all of the "other" elements outside the set) and remove that count from the "total". However, remember that we only have a Rule of *Sum* at our disposal, not a rule of Subtraction, so we should always be careful (for now, at least) to phrase such a step in terms of a *partition* and a *sum*. After that, we can subtract numbers or algebraic variables. Eventually, once we are more mathematically mature, we can easily skip this formality and just talk about "subtracting out" a count; for now, though, we want to emphasize the underpinnings of these counting arguments, so we will require this careful phrasing and application of the Rule of *Sum*.

(4) How many license plates have no vowels and no even digits?

This condition just limits the number of choices at each step. There are only 21 letters and 5 digits to choose from, so we get

$$26^6 + 26^7$$

total outcomes, by ROP and ROS.

8.2.3 Fundamental Counting Objects and Formulas

Let's return to our motivating example of counting poker hands. Remember that we want to know *how many* of each type of hand there are, how many *ways* we could be dealt a flush, say, from a freshly-shuffled deck of 52 cards. Let's start by answering a related, but simpler, question: how many *total* poker hands are there? Another way of phrasing the question–one that will actually hint at our method of answering it–is as follows: how many ways are there to shuffle the entire deck of 52 cards, and how many of those yield the same poker hand among the top 5 cards? That is, let's identify how many distinct (i.e. totally different) ways there are to shuffle the deck; let's call these ways *shufflings*. Then, let's think of a specific hand, say $T\clubsuit$ $J\clubsuit$ $Q\clubsuit$ $K\clubsuit$ $A\clubsuit$, and count how many shufflings have the property that the top 5 cards of the deck comprise that specific hand in *any* order (because we don't care *how* we receive the 5 card we're dealt, we just care what we're holding!).

What do we have at our disposal? That's right, the Rules of Sum and Product. That's pretty much it, other than our mathematical wit and intuition, so let's dive right in. How does shuffling a deck of cards correspond to a partition, or a multi-step process? Well, the interesting thing is that we don't actually care *how* the deck is shuffled, we only care about the number of *outcomes* of the process. What actually matters about a deck of cards? Right, the *order* of the cards from top to bottom. With that in mind, let's think about *constructing* an arbitrary shuffling by assigning the order of the cards.

Let's create a shuffling by taking a deck of cards in our hands and, one by one, placing a card face down on a stack in front of us. At the first step, we have 52 cards in our hands and no stack, so we have 52 choices. At the second step, we have 51 cards remaining in our hands to choose from, no matter what that first card was. (Remember: this is the important part of the Rule of Product, that the *number* of choices is independent of the actual choices made.) In the third step, we have 50 cards remaining, and so on. Eventually, in the 52nd step, we have only 1 card in our hands to place on the stack of 51 cards on the table. After that step is completed, we have a shuffling of the deck sitting in front of us, with the cards stacked face-down. The card from the 1st step is on the bottom, and the card from the last step is on the top. Moreover, we see that for any arbitrary shuffling, there is *exactly one* sequence of choices that produces that shuffling. (This satisfies that other part of the Rule of Product about having distinct outcomes. Think about this carefully and why it's required.)

These observations allow us to directly cite the Rule of Product to answer the question: how many shufflings of a standard deck of cards are there? The number is ...

$$52 \cdot 51 \cdot 50 \cdots 3 \cdot 2 \cdot 1 = \prod_{k \in [52]} k = 8.06581752 \times 10^{67}$$

Yowza! That's a big number. For the sake of comparison, Avogadro's Constant (the number of atoms in a mole) is on the order of 10^{23}. There is a much better notation for this kind of product that says "multiply all the natural numbers from 52 down to 1", and you've probably seen it before, but we'll define it now.

Definition 8.2.12. *Let $n \in \mathbb{N}$. The natural number $n!$, read as n* **factorial***, is given by*

$$n! = \prod_{k \in [n]} k = k \cdot (k-1) \cdot (k-2) \cdots 3 \cdot 2 \cdot 1$$

By definition, $0! = 1$.

(Recall that we used computing factorials as an example of applying the principle of induction to recursive programming, way back in Section 2.5.1. Read that section again!)

Let's think about what we've accomplished, in fact. What was special about the number 52 in this case? Besides it being the number of cards in our deck, nothing! What if we had posed the question: how many ways are there to put the elements of $[n]$ into an ordered list? If we replace n with 52, this is actually

the same questions as before! (We could just come up with a natural bijection between the set of cards and the set [52]. Can you do this? Do you see why this shows the questions are *equivalent*?)

Permutations

This type of question—how many ways are there to arrange n objects into an ordered list—is so common that we have a specific term for these ordered lists. We define them rigorously in terms of *functions*, but note their relationship to other mathematical objects (ordered list, for instance).

Definition 8.2.13. *Let $n \in \mathbb{N}$. A* **permutation** *of $[n]$ is a function $f : [n] \to [n]$ that is a bijection.*

Equivalently, a permuation of $[n]$ is an ordered n-tuple of elements from $[n]$ such that every element appears exactly once.

Proposition 8.2.14. *Let $n \in \mathbb{N}$. Let S be the set of all permuations on $[n]$. Then $|S| = n!$.*

Proof. We construct an arbitrary permutation of $[n]$ by selecting which element appears first in the ordered list. There are n options. Then, from all the elements *except* that one already chosen, select one to appear second in the list. There are $n-1$ options. In general, at step k, we choose from the $n-(k-1) = n-k+1$ elements not already chosen and pick one to appear next. This goes until step $n-1$, where we only have 1 option. By ROP, there are $n(n-1)(n-2)\cdots 2\cdot 1 = n!$ total outcomes. □

(Note: this motivates the convention of choosing to define 0! as 1. Since $n!$ represents the number of ways to permute n objects, and there is exactly 1 way to permute all of the elements of the empty set—there, we just did it!—it makes sense that $0! = 1$. This idea will return when we define *binomial coefficients* shortly; it will be very helpful to have $0! = 1$ for the corresponding formula.)

Selections

This mathematically proves a general version of our observation about shuffling cards, and it brings us closer to answering our original question about ranking poker hands. Remember that we hope to identify how many distinct shufflings of the deck yield a certain type of five card hand among the top five cards, so let's attack a slightly more general problem, first. Think of a *specific* five card hand, five particular cards. We're thinking of $T\clubsuit$ $J\clubsuit$ $Q\clubsuit$ $K\clubsuit$ $A\clubsuit$, so let's use that. Now, let's count how many deck shufflings place this specific hand among the top five cards.

How could we have such a situation? We don't care about the order in which we receive the cards in our hand, and we don't care about the order of the other 47 cards in the deck. All that matters is whether those specific cards are on the top. So let's follow the same idea we used before and *construct* a shuffling

with this property. We want to use the Rule of Product, so we need to identify a particular process that constructs a shuffling with the desired property. How can we do this?

There are really only two properties we need to satisfy, so let's identify a two step process that ensures those properties hold. The first step should place the 47 cards not from our hand on the bottom of the deck in some order. The second step should place the five cards from our hand on top of that pile in some order. The Rule of Product applies because no matter how we shuffle the bottom 47 cards, this doesn't affect the number of ways we can shuffle the top five cards. (In general, be careful to note *why* the Rule of Product applies in a given situation before applying it; this is often subtle and not obvious!) Now, we just need to count the number of ways to perform each step.

The first step involves creating a permutation of 47 cards. Proposition 8.2.14 tells us there are 47! ways to do this. The second step involves creating a permutation of five cards. Proposition 8.2.14 tells us there are 5! ways to do this. Then, the Rule of Product tells us the number of ways to complete these steps in succession is $47! \cdot 5!$. That's it!

What was special about our choice of $T\clubsuit$ $J\clubsuit$ $Q\clubsuit$ $K\clubsuit$ $A\clubsuit$ in this case? That's right, nothing! By applying the Rule of Product again, this fact will tell us something more about the number of shufflings of the deck. Specifically, let's say X is the number of ways to select a set of five cards as a poker hand. Now, consider the three step process of taking five particular cards from the deck, arranging them in some order, and then arranging the other 47 cards below it. The Rule of Product applies here because the number of ways to perform each step doesn't depend on the choices made in the previous steps. Furthermore, *every* shuffling of the deck arises from exactly *one* particular instance of this procedure. (Think about why this is true. Consider an arbitrary shuffling of the deck. The top five cards determine which hand we chose in the first step, the order of them determines how the second step was performed, and the order of the others determines how the third step was performed.) Thus, we have found two particular formulas for counting the same set of objects–that is, the shufflings of a deck of cards–and so it must be true that

$$X \cdot 5! \cdot 47! = 52!$$

and therefore

$$X = \frac{52!}{5! \cdot 47!}$$

Think about what this formula tells us. We let X designate the number of ways to choose a set of five cards from a set of 52 cards. What was special about five or 52? Again, that's right, nothing! We have essentially derived a formula for the number of ways to select any number of objects from a larger set of objects. It might not seem like it, but we are now very close to solving the poker hands problem. Before we finish that project, let's make one comment.

First, the type of argument we just made is a common and extremely useful proof technique in combinatorics. It is known as *counting in two ways*. What

we did was identify a particular set of objects–in this case, the set of shufflings of a deck of cards–and then describe two different procedures that allowed us to count the size of that set. Each procedure led to a different formula, and because we were counting the same set of objects, we know those formulas are equal. We will explore this type of argument more explicitly and see many examples in Section 8.4. For now, we hope that you can see why it is a valid argument type, especially because we will expect you to use it to prove Proposition 8.2.16 below! In doing so, you will be generalizing the argument we presented here. For illustration's sake, let's summarize what we did:

Argument Summary: We seek an expression for the number of ways to draw 5 cards from a deck of 52 cards. Let N be this number we are looking for. We will identify two different formulas for expressions that involve N. This will allow us to solve these algebraic expressions for a formula for N.

(1) Select an arbitrary and fixed five card hand. We will identify the number of ways to shuffle a deck of cards such that the top five cards are that fixed five card hand, in any order.

    Note that there are N ways to do this step. We seek a formula for N.

(2) Count the number of permutations of the entire deck of 52 cards.

(3) Count the number of permutations of the deck that yield those fixed five cards on the top. This is split into three steps:

    (i) Count the number of ways to permute those five cards.

    (ii) Count the number of ways to permute the other 47 cards.

    (iii) Count the number of ways to put those 5 permuted cards on top of those 47 permuted cards. (Note: There is only one way to do this, but it's important to point out as a separate step.)

(4) Overall, notice that we have counted the number of permutations (i.e. shufflings) of the deck in two separate ways, so they must be the same number.

(5) Simplify the expression (which involves N) to find a formula for N.

Now, let's generalize the formula we just derived. First, we make a definition and introduce some notation, and then we state a formula.

Definition 8.2.15. *Let $k, n \in \mathbb{N}$ with $n \geq k$. A* **k-selection** *from $[n]$ is an unordered set of k elements from $[n]$.*

The number of k-selections from $[n]$ is represented by $\binom{n}{k}$. This is known as a **binomial coefficient***, and is read as "n choose k".*

Proposition 8.2.16. *Let $k, n \in \mathbb{N}$ with $n \geq k$. The number of k-selections from $[n]$ is given by*

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Proof. Left for the reader as Exercise 2 in Section 8.2.4 □

Binomial Coefficients

One thing you might find surprising about the above formula is that the fraction is actually a natural number, no matter what k and n are! This is proven by the fact that it represents a number of ways to complete a procedure, as described in the proof, and this must be a natural number.

We want to point out one special case of this formula which may not occur to you. What if $k = 0$, say? What number should $\binom{n}{0}$ be? You might be surprised to find out that $\binom{n}{0} = 1$. Why does this make sense? Intuitively, we think of $\binom{n}{k}$ as the number of ways to select k objects from a set of n objects; so, how many ways can we select 0 objects from, say, 3 objects? Put 3 pens on your desk. Now, select none of them. There! You just did it! That was one way—and the *only* way—to select none of the objects. This argument works just as well when $n = 0$, even! Put no pens on your desk. Now, select none of them. There! You just did it in one way again. Thus,

$$\forall n \in \mathbb{N} \cup \{0\}. \binom{n}{0} = 1$$

There are "better", more mathematical reasons for this result, and we will point these out in the next section when we prove Pascal's Identity. For now, we hope that this heuristic explanation with selections makes sense and can convince you of this result.

Another fact is that $\binom{n}{K} = 0$ whenever $K > n$. This is because there are *no* ways to choose, for instance, 5 objects from a set of only 3 objects. This fact is borne out by our derivation above, because in one of the steps, we would be trying to (impossibly) draw more cards for a hand than there are cards *in* that deck, and there are 0 ways to do this. Then, when we apply ROP, the product would evaluate to 0.

If you play around with some values of k and n, you'll notice that the values of $\binom{n}{k}$ obeys a so-called **unimodal distribution**. That is, if we fix n and let k increase from 0 to n, we find the numbers going up, reaching a peak at $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$ (notice these are the same if n is even) and then decreasing again. Furthermore, the distribution is *symmetric* around that middle! Can you prove that these properties hold? Try it!

Arrangements

We now have all of the tools necessary to count poker hands (and plenty of other objects, for that matter). We know how many ways there are to permute the elements of a set, and we know how many ways there are to choose a subset of a certain size from a larger set. Between these two tools, we know how to count any combinations of cards. For instance, to count an *ordered* subset of cards, we can count the number of ways to choose the subset and *then* permute its elements, applying the Rule of Product to this two-step process. In fact, this idea is common enough that we will give it a defined name.

Definition 8.2.17. *Let $k, n \in \mathbb{N}$ with $n \geq k$. A* **k-arrangement** *from $[n]$ is an ordered k-tuple of elements from $[n]$ with no repeated elements.*

Equivalently, a k-arrangement from $[n]$ is a function $f : [k] \to [n]$ that is an injection.

Proposition 8.2.18. *Let $k, n \in \mathbb{N}$ with $n \geq k$. The number of k-arrangements from $[n]$ is given by $\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!}$.*

Proof. Left for the reader as Exercise 3 in Section 8.2.4 □

Repetition

Before we go on and count those poker hands, actually, we should point out that all of the standard counting formulas we have seen in this section only consider procedures where objects are not allowed to be *repeated*. That is, when we choose a five card hand from a deck, we can't have two $A\clubsuit$s, for instance. There are situations where we will want to allow objects to be selected multiple times. Look back at the License Plates example in the previous section. We were allowed to repeat any digit/letter; for instance, 111AAA is a valid license plate. Let's see one more example here:

Example 8.2.19. Consider a standard, fair, two-sided coin. Flip the coin 6 times in a row and write down the outcomes, either H or T for each flip.

Question: How many possible sequences of outcomes are there?

To answer this question, we note that there are 2 possible outcomes on each flip, regardless of the outcomes on the previous flips. Thus, the Rule of Product applies, and we can say there are $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$ possible sequences of flips.

The reason this idea is related to selections and arrangements (beside using the Rule of Product, of course) is that we can also represent these sequences as arrangements of 4 objects from the set $\{H, T\}$ where objects are allowed to appear more than once. (There is a natural correspondence between $\{H, T\}$ and $[2]$, so it is like we are arranging 4 objects from $[2]$, where the objects can occur more than once.)

This general idea is conveyed by this definition:

Definition 8.2.20. *Let $k, n \in \mathbb{N}$. A* **k-arrangement with repetition** *from $[n]$ is a k-tuple of elements from $[n]$ where elements are allowed to appear more than once.*

Notice that there is *no restriction* on k because we are allowing elements to appear multiple times. Before, with k-arrangements without repetition, it wouldn't make sense to choose 10 objects from 8 objects if we couldn't repeat any! Here, though, this is allowed, so k and n can be *any* natural numbers.

Proposition 8.2.21. *Let $k, n \in \mathbb{N}$. The number of k-arrangements with repetition from $[n]$ is given by n^k.*

Proof. Left for the reader as Exercise 4 in Section 8.2.4. □

You might anticipate a definition and proposition for *k-selections* with repetition that are similar to the ones for arrangements with repetition. We will discuss these in Section 8.5, but the techniques used to count them are more advanced than the ones we have now, so we will address this later.

Summarizing Counting Formulas

Let's summarize the standard counting objects and formulas we have defined and derived thus far: Say we have n objects and we want to select k of them. How many ways can we do this? The answer depends on *two questions:*

- Are repeats allowed?
- Does order distinguish the outcome?

Each of these questions can be answer with Yes or No, and each of the four ways to answer them yields a different formulation of the original question.

Order Matters? \ Repeats?	Yes	No
Yes	n^k	$\dfrac{n!}{(n-k)!}$
No	???	$\binom{n}{k}$

(Note: Sometimes, the roles of n and k are reversed in a problem. Be careful about this! We'll try to stick to these conventions but, in general, the letters aren't important; it's what they *represent*.)

Combinatorics Definitions in terms of Functions

Remember there are also equivalent formulations of these counting ideas in terms of *functions*, and it's helpful to have this in mind. Perhaps representing a problem in terms of functions will help us solve it. At the very least, it's a good mental exercise to work through and make sure you understand the relationship between, for instance, *permutations* and *bijections*. We will just state each of these formulations (and some corresponding formulas) and ask you to think about them on your own. Try to see exactly why and how the notions are related; try to explain them to a friend who only knows one of the interpretations; work with your classmates to perhaps come up with a different formulation!

- A **permutation** of n elements is a bijection $f : [n] \to [n]$.

  There are $n!$ possible bijections from the set $[n]$ to itself.

- An **arrangement** of k elements from n elements is an injection $f : [k] \to [n]$.

  There are $\frac{n!}{(n-k)!}$ injections from $[k]$ to $[n]$.

- An **arrangement with repetition** of k elements from n elements is a function $f : [k] \to [n]$.

  There are n^k possible functions from $[k]$ to $[n]$.

8.2.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between a **selection** and an **arrangement**?

(2) How might a **permutation** be defined in terms of selections and arrangements?

(3) What is $\binom{10}{15}$?

(4) How is a permutation related to the concept of a **bijection**?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Verify algebraically that $\binom{n}{k} = \binom{n}{n-k}$.

(2) Prove Proposition 8.2.16, i.e. prove that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Do this by adapting the argument we used for counting the number of 5 card hands from a standard deck.

(3) Prove Proposition 8.2.18. That is, prove there are

$$\frac{n!}{(n-k)!}$$

possible k-arrangements from $[n]$.

(4) Prove Proposition 8.2.21. That is, prove there are

$$n^k$$

possible k-arrangements with repetition from $[n]$.

8.3 Counting Arguments

Now we are fully ready to address the motivating problem of this chapter! We will employ the counting techniques we have developed–the Rules of Product and Sum–as well as the formulas for selections and arrangements. Importantly, we will show you some standard counting arguments and proof strategies. We will point out some general guidelines and proof techniques as we go, motivating and implementing these with several examples. These are the types of techniques we will expect you to use in the future.

8.3.1 Poker Hands

Example 8.3.1. **One Pair**

Let's start near the bottom of the ranks and count the number of poker hands that correspond to *one pair*. We emphasize that we only want to count hands with *exactly* one pair, and exclude two pairs, three of a kinds, full houses and four of a kinds. This idea will surface soon enough in our counting argument. (It also hints at why counting "high card" hands is actually quite difficult, far more intricate than just selecting five random cards! How can we guarantee that a hand has *no* matching cards, isn't a straight, and isn't a flush? We will address this question later in this section.)

In this example–and in every other example we will explicate here, and every other exercise you will complete (do you get the sense this is important?)–we seek a *process* wherein we *construct* an object (in this case, a poker hand) with the desired properties (in this case, having exactly one pair and no other matching cards). By counting the number of options at each step in the process, and ensuring that every desired object can only be obtained via one set of options in the process, we can apply the Rule of Product and identify the number of objects with the desired properties!

Here's a useful strategy for coming up with these processes: pretend your friend is holding one of the objects you're counting in his/her hands, but you can't see it. What questions would you ask to identify the particular propeties of the object he/she is holding? These can be yes/no questions or, more often than not, queries about the particular properties the object has. In our specific case, counting one-pair hands, we would likely ask the following questions: (1) "What are the two cards in the pair?" and (2) "What are the three cards not in the pair?" With the answers to those questions, we could fully specify the hand our friend is holding. Unfortunately, it's too hard to count the number of answers to those questions as they are posed. We should be more specific and

break our questions into smaller parts. That way, we can count the number of answers to each question and use those numbers in the Rule of Product.

How can we be more specific? How can we break question one into parts? Imagine the types of answers our friend might give us for question one. We might hear something like, "The Ace of Hearts and Ace of Spades" or "The Sevens of Diamonds and Clubs". This signals the important properties of an answer to question one: we need to know the *rank* of the pair cards (are they both Aces? Kings? Queens? etc.) and the *two suits* represented. We know there are 13 ranks and 4 suits in the deck. With this information, we can identify how to *construct* a pair and count the options.

1. Choose a rank for the two cards in the pair: 13 options
2. Choose the two suits for those cards: $\binom{4}{2} = 6$ options

Notice that we have used the binomial coefficient $\binom{4}{2}$ to signify that we are selecting 2 suits from a set of 4 suits, so there are $\binom{4}{2}$ ways to do this.

Note: $\binom{4}{2}$ is a NUMBER. It represent the number of *ways* to do something, and does *not* actually correspond to doing that action. That is, we don't say something silly like "$\binom{4}{2}$ selects 2 suits from the set of 4 suits." How can a number choose cards from a deck?

Also note: We wrote $\binom{4}{2} = 6$ in this case for illustration's sake but, in general, we do *not* expect (or even necessarily want) you to evaluate binomial coefficients. The arithmetic often involves very large numbers and, quite frankly, the number $\binom{4}{2}$ is far more illustrative than 6. It indicates to a reader that this step in your process involves selecting 2 elements from a set of 4, whereas 6 could represent $\binom{6}{1}$ or $2 \cdot \binom{3}{2}$ and so on. With that observation made, we might as well write the number in the first step as $\binom{13}{1}$, right?

Now, we observe that any selections made in these steps produce a *unique* pair. That is, we can't possibly have a pair that could arise from two different versions of this proces. Thus, the Rule of Product applies, and we can conclude that there are $\binom{13}{1} \cdot \binom{4}{2}$ ways to select a pair of cards.

What if we had performed these two steps in the opposite order? We could just as well identify a pair of cards by asking which two suits are represented and *then* asking what their common rank is? (Of course, this only works if we know, a priori, that the cards have a common rank.) In that case, the Rule of Product would tell us there are $\binom{4}{2} \cdot \binom{13}{1}$ such pairs. Hey, that's the same number! The commutativity of multiplication of real numbers (that is, $x \cdot y = y \cdot x$ for any $x, y \in \mathbb{R}$) confirms our intuition that these steps are reversible.

We aren't quite done constructing a *poker hand* with one pair. We need to choose three more cards. What property should they have? What more specific questions could we ask our friend, besides "What are they?". We need to know the three cards' *ranks* and their *suits*. Is there any restriction on their suits? No! (Because we have a pair already, there is no chance for a flush.) Is there any restriction on their ranks? Yes! We know the three cards all have different ranks, and none of them match the rank of the pair already chosen. With these observations, we can reverse the process and *construct* the rest of the hand.

1. Choose 3 ranks from the 12 remaining (i.e. not the same rank as the pair cards): $\binom{12}{3}$ options
2. Arrange those 3 ranks in increasing order: 1 option
3. Choose a suit for the lowest-ranked card: $\binom{4}{1}$ options
4. Choose a suit for the middle-ranked card: $\binom{4}{1}$ options
5. Choose a suit for the highest-ranked card: $\binom{4}{1}$ options

Why did we need step 2? Look back at the definition of *selection*; it is an *unordered* list, or a *set*. Thus, it wouldn't make sense to jump into step 2 by saying "Choose a suit for the 1st of those chosen cards" because, well, there is no *1st* card! We need to *impose* some kind of ordering on the cards to refer to them individually. You might be tempted to order them as we remove them from the deck. This would break step 1 into 3 sub-steps: (a) choose the 1st card: $\binom{12}{1}$ options; (b) choose the 2nd card: $\binom{11}{1}$ options; (c) choose the 3rd card: $\binom{10}{1}$ options. Applying the Rule of Product to this step yields a *different* number than step 1:

$$\binom{12}{1} \cdot \binom{11}{1} \cdot \binom{10}{1} = 12 \cdot 11 \cdot 10 \neq \binom{12}{3} = \frac{12!}{3! \cdot 9!} = \frac{12 \cdot 11 \cdot 10}{6}$$

This is because the (a)-(b)-(c) step imposes an order on those three cards that doesn't actually matter within our poker hand. When playing cards, you don't care *how* you receive your cards, only what they are! (However, notice that if we "divide out" by the number of ways to order 3 cards, namely 3!, we get the same number. This hints at an interesting concept, a kind of "inverse" of the Rule of Product. We will discuss this at the end of this section.) This is why we couldn't refer to "the 1st card" in step 2. Instead, we found an *inherent* ordering of the cards, a particular property they possess that allows us to refer to specific cards among them without applying an external ordering.

Again, the Rule of Product applies because any selection of 3 cards of different ranks could only come from one set of choices made in these steps. Furthermore, we can think of selecting a pair as Step 1 and selecting three other cards of different ranks as Step 2 and apply the Rule of Product to this *entire* process. This finally gives us an answer for the number of "one pair" poker hands:

$$\binom{13}{1} \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot \binom{4}{1}^3$$

Notice that we have combined the three numbers from the last steps above into one coefficient raised to the third power. Now, this type of numerical answer is *totally acceptable* and is far better than just writing down $1,098,240$. If you make a "typo" on your homework or make a calculator error, how can we identify the error and offer a comment? ☺

We did previously note the commutativity of multiplication and the idea of doing steps in different orders. However, we hope you'll agree that explaining a product like

$$\binom{4}{1}^3 \cdot \binom{13}{1} \cdot \binom{12}{3} \cdot \binom{4}{2}$$

even though it represents the same process, is far more intricate, and unnecessarily so, at that.

We chose to be particularly wordy with our explanation in the last subsection. We won't expect you to write *nearly* as much. We were just officially introducing a formal method that applies the counting rules and formulas we developed in the last section, while also mentioning some heuristic rules and strategies to approach problems. So with that said, let's present a typical solution to this problem, in more condensed form. This is the type of solution we will expect you to write:

Question: How many 5-card poker hands are "one pair" hands?

Answer: We claim there are

$$\binom{13}{1} \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot \binom{4}{1}^3$$

such hands. To show this, we will identify a four-step process and apply ROP. An outcome of this process is a "one pair" poker hand:

(1) Select a rank to constitute the pair.

There are $\binom{13}{1}$ ways to do this.

(2) Select which two suits of that card in (1) appear in the hand.

There are $\binom{4}{2}$ ways to do this.

(3) Select three *other* ranks to appear.

There are $\binom{12}{1}$ ways to do this.

(4) For each rank chosen in (3), select a suit of that card to appear in the hand.

There are $\binom{4}{1}$ ways to do this, each of three times; thus, there are $\binom{4}{1}^3$ ways in total.

Applying ROP, we find the answer given above. □

Does this make sense? Notice how much shorter it is than our explanation above. This is fine! We will continue to sometimes write out some details in our written examples here (to help you understand how to *approach* these problems, before writing them up), but your written solutions can be a little more condensed, as long as they identify all the key elements of the problem's solution. Notice that we pointed out a use of the ROP, cited it, and identified all the steps in the process; for each step, we noted how many ways there are to do that step. It just so happens each of these steps are pretty simple, and

the number of ways to perform them is clear in each case. In general, we might expect a more thorough description. For instance, we would consider writing that the number of ways to do step (3) is $\binom{12}{1}$ because we aren't allowed to re-select the rank chosen in step (1). However, we felt this was clear from the descriptions so we left it out. This is a judgment call, though, and we recommend (as always) setting aside your proofs and rereading them as if you didn't write them. If you can't remember, or aren't entirely sure, why something is true, consider adding a little extra description there.

Before doing another example, let's point out a *different* solution to this same problem!

Question: How many 5-card poker hands are "one pair" hands?

Answer: We claim there are

$$\binom{13}{4}\binom{4}{1}\binom{4}{2}\binom{4}{1}^3$$

"one pair" poker hands. We will identify a six-step process and apply ROP. The main idea is that a one pair hand can be identified by choosing all four ranks that appear and identifying which one is repeated twice (leaving the others to appear just once).

(1) Select 4 ranks that will appear in our hand.

There are $\binom{13}{4}$ ways to do this.

(2) Of the 4 ranks selected in Step (1), select one of them. Two cards of that rank will appear in our hand.

There are $\binom{4}{1}$ ways to do this.

(3) For that rank chosen in Step (2), select 2 suits. These will appear in our hand. There are $\binom{4}{2}$ ways to do this.

(4) For the lowest of those 3 ranks *not* chosen in Step (2), select a suit.

There are $\binom{4}{1}$ ways to do this.

(5) For the middle of those 3 ranks *not* chosen in Step (2), select a suit.

There are $\binom{4}{1}$ ways to do this.

(6) For the highest of those 3 ranks *not* chosen in Step (2), select a suit.

There are $\binom{4}{1}$ ways to do this.

By ROP, and simplifying $\binom{4}{1}\binom{4}{1}\binom{4}{1} = \binom{4}{1}^3$, we have shown the expression above is correct. □

Isn't that neat? We'll leave it to you to verify that

$$\binom{13}{4}\binom{4}{1}\binom{4}{2}\binom{4}{1}^3 = 1098240 = \binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3$$

is true, numerically speaking. Without calculating that number in the middle, though, we could be *sure* that the two expressions, on the left and right, are absolutely *equal* representations of the same number because they count the *same* thing: the number of one pair poker hands. This is another instance of that idea of "counting in two ways" that we are building towards.

Example 8.3.2. **Flush**

Let's jump right into another problem and solve it. Let's count the number of poker hands that are flushes. A flush hand is defined by two propeties: the *suit* all 5 of its cards share, and the 5 ranks of those cards. Thus, a flush can be generated by a two-step process:

(1) Select a suit for all five cards of the hand.

There are $\binom{4}{1}$ ways to do this.

(2) Select five of the cards from that suit to appear in the hand.

There are $\binom{13}{5}$ ways to do this.

Since each flush hand is uniquely defined by these two steps, we can apply ROP and conclude that there are

$$\binom{4}{1} \cdot \binom{13}{5} = 5148$$

poker hands that are flushes.

This proof given in this example (except for the final number 5148, which we only included here for sake of comparison to the "one pair" answer which was *much* larger), is completely correct and rigorous, and would receive full credit. Use this as a model for simple counting problems with the Rule of Product.

Example 8.3.3. **Straight**

The ranks of the cards in a straight are uniquely determined by the "starting rank", the lowest card of the hand. If I told you I had a 5-card straight starting with 7, you'd know immediately I have a 789TJ straight. Since we can have a straight like A2345, or one like 23456, ... all the way up to TJQKA (Note: There is no "going around the corner" in a straight, like QKA23), this means we have 10 possible *lowest ranks* in a straight. Thus, there are ten types of straight, and after picking which type we have, we just need to assign the suits so that they aren't all the same (in which case we'd have a straight flush).

We claim there are

$$\binom{10}{1}\left[\binom{4}{1}^5 - \binom{4}{1}\right] = 10 \cdot (4^5 - 4)$$

5 card hands that are straights.

Proof. We will describe 5 card hands that are straights by a two-step process:

1. Select one of 10 ranks to be the *lowest* rank in the straight. These options are A,2,3,4,5,6,7,8,9,T, so there are **10 options** in this step.

   Note: This *determines* the other 4 ranks in the hand, since the 5 ranks must be consecutive and we know what the lowest one is.

2. Assign suits to the 5 cards so that they are not *all* the same suit.

   Let's say X is the set of all possible ways to assign suits in this manner, so there are $|X|$ options in this step.

We will now find $|X|$ by establishing a partition. Let Y be the set of all assignments of 5 suits so that they *are* all the same. Notice that the sets X and Y form a partition of U, the set of all possible assignments of 5 suits. (That is, any assignment of 5 suits either selects all the same suit, or it does not.) Thus, by ROS, we have $|U| = |X| + |Y|$.

We can find $|U|$ by a 5 step process, where in Step i, we select one of the 4 suits for the i-th highest rank in the hand. With 4 options at each of 5 steps, we have $|U| = 4^5$.

We can find $|Y|$ by noticing that any such selection amounts to picking one of the 4 suits and assigning that suit to all 5 cards in the hand. Thus, $|Y| = 4$.

Accordingly, we can rearrange the above equality and write

$$|X| = |U| - |Y| = 4^5 - 4$$

Since $|X|$ is the number of options in Step (2) above, by ROP, we have proven the claim. □

Note: In this proof, we came up with all of the relevant steps to show that there are $10 \cdot 4 = 40$ possible *straight flushes* (straights of the same suit), only $1 \cdot 4 = 4$ of which are *royal straight flushes* (TJQKA of the same suit). Try to write out those arguments for yourself!

8.3.2 Other Card-Counting Examples

Let's look at some related examples to broaden the class of techniques we're applying.

Example 8.3.4. **At least 3 Aces**

For this example, let's count the number of poker hands that have at least three aces. Again, let's apply the technique we used above and think of the essential *properties* of such a hand. Try to think of a few questions yourself, with the goal being that the answers determine a *unique* hand and, given any answer, we can count exactly how many ways to construct a hand that yields that answer.

Did you notice the difficulty? One of the answers to the questions *directly affects* the nature of the other questions! This indicates some deeper mathematical issues at play. Perhaps it makes sense to have that determining question come *first*, and then consider what decisions must be made from there.

First, IF there are exactly 3 Aces in the hand, then we need to determine the characteristics of the other two cards. Those two cards are either (a) the same rank or (b) two different ranks. Thus, there are two sub-cases for this particular case. This yields the following procedure:

1. Choose 3 suits for the 3 Aces: $\binom{4}{3}$ options
   (a) The remaining two cards are of different ranks:
      i. Choose 2 ranks from the remaining 12 for the other 2 cards: $\binom{12}{2}$ options
      ii. Choose a suit for the lowest-rank card chosen in Step 2: $\binom{4}{1}$ options
      iii. Choose a suit for the highest-rank card chosen in Step 2: $\binom{4}{1}$ options
   (b) The remaining two cards are of the same rank:
      i. Choose 1 rank from the 12 non-Ace ranks: $\binom{12}{1}$ options
      ii. Choose 2 suits for this rank: $\binom{4}{2}$ options

Then, by the Rules of Product *and* Sum (since we have separate cases in a process), we find there are

$$\binom{4}{3}\left[\binom{12}{2}\binom{4}{1}^2+\binom{12}{1}\binom{4}{2}\right]$$

hands with *exactly 3 Aces*.

Second, IF there are exactly 4 Aces in the hand, we need to determine the characteristic of the fifth card in the hand. This yields the following procedure

1. Choose 4 suits for the 4 Aces: $\binom{4}{4}$ options
2. Choose 1 rank from the remaining 12 for the other card: $\binom{12}{1}$ options
3. Choose a suit for the card chosen in Step 2: $\binom{4}{1}$ options

We can apply the Rule of Product and conclude that there are then $\binom{4}{4}\binom{12}{1}\binom{4}{1}$ hands with *exactly 4 Aces*. Now, we must apply the Rule of Sum! What we have here is a *partition* of the set of desired hands–those with at least three Aces–into two subsets–those with exactly three Aces and those with exactly four Aces. Since those subsets partition the larger set (i.e. every hand with at least three Aces has either three Aces or four Aces, not both and not neither), we may apply the rule of sum and conclude that there are

$$\binom{4}{3}\left[\binom{12}{2}\binom{4}{1}^2+\binom{12}{1}\binom{4}{2}\right]+\binom{4}{4}\binom{12}{1}\binom{4}{1}$$

poker hands with at least three Aces.

Recall that the rigorous statement of the Rule of Sum concerned cardinalities of finite sets, and yet we didn't technically get into those details in the previous example. There is a certain amount of discretion and finesse required with these types of combinatorial arguments. Is it obvious to you that *every poker hand with at least three Aces has either exactly three or exactly four Aces, not both and not neither*? We are not saying it should be totally obvious and you're a dummy for not seeing it right away! Far from it! What we are saying is that this type of statement should probably suffice as an explanation in a proof. Yes, we could dive into further detail, reformulate poker hands in terms of sets, and completely rigorize the game of poker into set notation. What good would that really do, though? It seems far easier to explain it via the italicized statement above. If we were pressed for details by a confused reader, we could offer further explanation, but for a general audience, this argument would suffice. Hopefully, this rule of thumb–convincing a general audience, but being able to explain further when pressed further–should guide you into making decisions about how much detail to include in a counting argument. The essential observation here is that we indicated *why* our choices pertain to a partition of the set of hands in question. No, we didn't rigorously prove the two sets were disjoint, but we offered an explanation as to why.

Another approach to this problem does *not* involve considering the suits of the non-Ace cards. Instead, we can approach the process of constructing a poker hand with at least 3 Aces as follows:

1. If there are exactly 3 Aces:
   (a) Choose 3 suits for the 3 Aces: $\binom{4}{3}$ options
   (b) From the remaining 48 non-Ace cards, select 2 to "fill out" the 5 card hand: $\binom{48}{2}$
2. If there are exactly 4 Aces:
   (a) Choose 4 suits for the 4 Aces: $\binom{4}{4} = 1$ option
   (b) From the remaining 48 non-Ace cards, select 1 to "fill out" the 5 card hand: $\binom{48}{1}$

Thus, by the Rule of Sum (since we have partitioned the hands based on how many Aces they have) and by the Rule of Product in each of the two cases, we have

$$\binom{4}{3}\binom{48}{2} + \binom{4}{4}\binom{48}{1}$$

total poker hands with at least 3 Aces. You will see (and use) this approach more often. The previous argument was more similar to the previous example involving flushes, so that's what we presented first. This argument is a bit shorter and "slicker", and is thus more commonly used. But wait a minute, these answers look different! We were counting the same set of poker hands, so

shouldn't we expect the same *final number*? Well, yes, and we recommend that you perform the requisite algebraic manipulations to convince yourself that

$$\binom{4}{3}\binom{48}{2}+\binom{48}{1}=\binom{4}{3}\left[\binom{12}{2}\binom{4}{1}^2+\binom{12}{1}\binom{4}{2}\right]+\binom{4}{4}\binom{12}{1}\binom{4}{1}$$

It will only take a minute, and it is worthwhile.

Before moving on to another problem, let's look at a *false argument* about this one. It may seem strange to look at wrong answers, but we know from experience that it can be extremely helpful and instructive to try to *find the flaw* in a faulty argument. Sure, we could just compare two large integers and just say, "Hey, look, they're different!" but this is not enlightening. Rather, we want to follow a combinatorial argument and pinpoint the step that makes a logical flaw or alters the set of objects we are counting in a flawed way. We highly recommend this tecnique for several reasons. First, it gives you good practice with reading proofs and understanding others' arguments. This will help you as you learn more mathematics and read other books that might not explain things in exactly the same way. Second, it helps you become a better editor of your own proofs. After writing up a homework problem, set it aside for half an hour and come back to it with a fresh mind. Read it as if you didn't write it (as best you can, we understand you just can't pretend you didn't do it!). Does it make sense? Are there certain steps that seemed obvious when you wrote them but whose details escape you now? Is the answer even correct and are you convinced by it? Third, recognizing when a bad step is made in a proof solidifies your understanding of the principles underlying the argument. Going through combinatorics arguments and identifying flaws will really help your intuition and understanding of the Rules of Sum and Product. Trust us.

What do you make of this argument? Remember, this answer is *incorrect*, and we want to know why!

Example 8.3.5. **Find the Flaw!** How many 5-card poker hands have at least three Aces?

1. For hands that have three Aces:
   (a) Choose 3 of the 4 Aces: $\binom{4}{3}$ options
   (b) From the remaining 49 cards, choose 2 more: $\binom{49}{2}$ options
2. For hands that have four Aces:
   (a) Choose 4 of the 4 Aces: $\binom{4}{4}=1$ option
   (b) From the remainign 48 cards, choose 1 more: $\binom{48}{1}$ options

Thus, there are

$$\binom{4}{3}\binom{49}{2}+\binom{4}{4}\binom{48}{1}$$

poker hands with at least 3 Aces.

What's the problem here? Do you see any errors? Was the Rule of Product applied inappropriately? Was the Rule of Sum applied to something that isn't actually a partition? Did we overcount? Undercount? Did we count some hands that do not have the desired properties? Think about this before reading on.

Here's what we noticed: this answer is *too large*. We *overcounted* by including certain hands multiple times in our count. That is, every hand we sought to count is included at least once by the steps above, but some hands can be constructed in *multiple ways* via those steps. These observations guarantee that our number is too large.

How did we know this? We recommend actively trying to identify a hand that can be constructed in different ways by following the above steps. If you're reading through a proof and can do this, you know that the entire proof is now flawed. In this case, let's examine a hand that has exactly 4 Aces; specifically, let's look at the hand $A\clubsuit A\spadesuit A\diamondsuit A\heartsuit 2\clubsuit$. We can construct this hand by the following paths through the steps:

1. Choose 3 of the 4 Aces: $A\clubsuit A\spadesuit A\diamondsuit$
2. From the remaining 49 cards, choose 2 more: $A\heartsuit 2\clubsuit$

Or, we could take this path:

1. Choose 4 of the 4 Aces: $A\clubsuit A\spadesuit A\diamondsuit A\heartsuit$
2. From the remaining 48 cards, choose 1 more: $2\clubsuit$

Do you see the problem now? This exact same hand is produced in (at least) two distinct ways via the process outlined above. Thus, the answer is an overcount. Are there any other ways we could construct this same hand? How many? Try to identify another hand that is overcounted. Can we possibly identify how many times every hand is overcounted by and amend our answer that way? This is an interesting (and very challenging, actually!) idea that we'll return to later.

Potential Flaws in Arguments

For now, we want to emphasize the technique of reading combinatorial proofs and looking for some standard flaws:

- **Misuse of Rule of Product**:
  The proof incorrectly applies the Rule of Product to a situation that doesn't warrant it. Perhaps the number of options at each step of the procedure change somehow, depending on how the previous steps are completed. Or, perhaps different sequences of steps produce the same outcome.
- **Misuse of Rule of Sum**:
  The proof incorrectly applies the Rule of Sum to a situation that doesn't warrant it. Perhaps the sets of the "partition" are not actually disjoint.

Or, perhaps the union the sets of the "partition" do not actually cover the entire set in question.

- **Overcount**:
  Every desired object is counted at least once, but some are counted more than once. That is, some elements of the set in question can be counted in multiple ways via the steps of the proof.
- **Undercount**:
  Some desired objects are not counted at all. That is, some elements of the set in question are not counted by the steps of the proof.
- **Extraneous Count**:
  Some undesired objects are counted. That is, the steps of the proof count some objects that are not elements of the set in question.

We recommend reading over your written proofs and trying to identify these flaws, even if they aren't there. Perhaps by *struggling* to find an overcounting argument, say—by attempting to construct certain objects in multiple ways via your steps—you actually identify a flaw you didn't know was there! If you don't find any flaws, you can be more assured that your proof is fully correct.

Example 8.3.6. Here is a standard example of a naive **overcount**. We will show how it is an overcount and then fix it by counting in a different way! Here is the question:

How many 5-card hands have at least one card of each suit?

Here is an **incorrect** argument:

> There are $\binom{13}{1}^4 \cdot \binom{48}{1} = 1370928$ such hands.
>
> We can use a five-step process. In step 1, we select one of the 13 Hearts. In step 2, we select one of the 13 Diamonds. In step 3, we select one of the 13 Spades. In step 4, we select one of the 13 Clubs. There are $\binom{13}{1}$ ways to do each of these steps.
>
> Next, from the remaining 48 cards, we select one of them to complete our 5-card hand. By ROP, the claim above follows.

What's wrong with this? Think about it carefully before reading on. Look at the list of potential mistakes above; does one of them apply here? How would you show this?

We think that this is an **overcount**. To show this, we will exhibit a particular 5-card hand that should be counted only once but is, in fact, counted *at least twice* by the procedure outlined in the argument above.

Consider the hand $A\heartsuit, A\diamondsuit, A\spadesuit, A\clubsuit, K\heartsuit$. Notice that this hand can be achieved by the above procedure in two ways:

(1) Step 1: pick $A\heartsuit$. Step 2: pick $A\diamondsuit$. Step 3: pick $A\spadesuit$. Step 4: pick $A\clubsuit$. Step 5: pick $K\heartsuit$.

(2) Step 1: pick $K\heartsuit$. Step 2: pick $A\diamondsuit$. Step 3: pick $A\spadesuit$. Step 4: pick $A\clubsuit$. Step 5: pick $A\heartsuit$.

Since a hand is *unordered*, these two procedures yield the *same outcome*. However, the argument above would count these two outcomes separately. Thus, the argument is an overcount.

To fix this argument, let's think more carefully about **how many** of each suit must appear. With 5 cards to be had, and only 4 suits, we see that requiring at least one of each suit means we have three suits that appear once and one suit that appears twice. That is the *only* way this could happen. Stated another way, the *distribution* of the suits has to look like $(1, 1, 1, 2)$.

To count the number of such hands, we identify a process:

- Select which of the four suits will appear twice. (The other three are fixed to appear once each.)

  There are $\binom{4}{1}$ ways to do this.

- From that suit, select two cards.

  There are $\binom{13}{2}$ ways to do this.

- From each of the other three suits, select one card.

  There are $\binom{13}{1}^3$ ways to do this.

By ROP, we find there are

$$\binom{4}{1}\binom{13}{2}\binom{13}{1}^3 = 685464$$

many 5-card hands with at least one card of each suit.

Example 8.3.7. **At most 2 Aces**

Let's pose a similar problem now. How many 5-card poker hands have *at most* 2 Aces? Try this one on your own for a few minutes before reading on. If you're struggling, try to follow a similar argument to the one we made in the last problem. What are the similarities and differences of these two problems?

Here's how we handled this problem:

1. For hands with exactly 2 Aces:
   (a) Select 2 of the 4 Aces: $\binom{4}{2}$ options
   (b) From the 48 remaining non-Aces, select 3: $\binom{48}{3}$ options
2. For hands with exactly 1 Aces:

(a) Select 1 of the 4 Aces: $\binom{4}{1}$ options

(b) From the 48 remaining non-Aces, select 4: $\binom{48}{4}$ options

3. For hands with exactly 0 Aces:

(a) Select 0 of the 4 Aces: $\binom{4}{0} = 1$ options

(b) From the 48 remaining non-Aces, select 5: $\binom{48}{5}$ options

Since cases 1 and 2 and 3 don't overlap (i.e. a poker hand has a specific number of Aces), we may apply the Rule of Sum; also, we may apply the Rule of Product in each of the three cases because we perform the two steps in succession. Thus, there are

$$\binom{4}{2}\binom{48}{3} + \binom{4}{1}\binom{48}{4} + \binom{48}{5}$$

(Note: it is common to omit the $\cdot$ multiplication symbol between binomial coefficients; the multiplication is implicitly assumed.)

Did you remember to count hands with 0 Aces? Forgetting this case is a common mistake! Did you also avoid the overcounting argument we saw in the last problem? We need to partition the set of hands in question by identifying three non-overlapping cases, depending on how many Aces are in the hand.

Another perfectly reasonable approach to this problem is to take advantage of the work that we've already done in the previous example. Perhaps you thought of this approach? If so, kudos for your cleverness! The main idea is to partition the set of *all* poker hands into two distinct cases. Every poker hand must either have at most 2 Aces *or* at least 3 Aces. Right? Let's let S be the set of all poker hands with at most 2 Aces, T be the set of poker hands with at least 3 Aces, and H be the set of all poker hands. Our explanation says that $H = S \cup T$ and $S \cap T = \varnothing$. Thus, the Rule of Sum can be applied to deduce that $|H| = |S| + |T|$. Furthermore, since we need to identify $|S|$, we can write this as

$$|S| = |H| - |T|$$

and therefore

$$|S| = \binom{52}{5} - \left(\binom{4}{3}\binom{49}{2} + \binom{4}{4}\binom{48}{1}\right)$$

We were able to write down this solution without counting anything else! All we needed was that partition into two sets whose cardinalities were already *known*.

This strategy indicates a deeper principle at play. In essence, we applied the "Rule of Subtraction" to find the answer we cared about. This amounted to applying the Rule of Sum, as it was stated previously, and then manipualting an expression from there. Indeed, this is the "right" way to think about it, in the sense that this is the way the underlying mathematical principles are applied. However, it is common to see the "Rule of Subtraction" applied more directly, in a way, in mathematical proofs. A proof-writer might assume some familiarity, on the part of the reader, with the sophisticated workings of the Rule of Sum and "jump" to a conclusion without explicitly identifying a partition or

rigorously explaining how the Rule of Sum was applied. For instance, a higher level mathematician might offer a proof to this current example by writing the following:

> From the set of all poker hands, remove those that have three or four Aces, yielding
>
> $$\binom{52}{5} - \binom{4}{3}\binom{48}{2} - \binom{48}{1}$$

A fellow mathematician, after a moment's thought, would accept this proof. However, we agree with what you might be thinking: isn't that too *short*? Doesn't it make the reader think too hard? For now, at this point in your mathematical career, we strongly encourage (and **require**) you to provide more *explicit* details in a proof like this. We expect you to apply the Rule of Sum and indicate why there is a *partition* underlying that application, and then manipulate any algebraic expressions to draw a conclusion. Later on, outside of this course, feel free to use the "Rule of Subtraction" as you see fit. For now, though, we want you to get a proper handle on the underlying principles, and that is why the Rule of Sum is required.

Here is one final hand-counting question. It involves both the Rules of Sum and Product, and requires some careful thinking about the steps of your process

Example 8.3.8. **Exactly 1 Queen and exactly 1 ♠**

How many poker hands have exactly one Queen and exactly one Spade?

Try this on your own for a little while. Think about asking questions of your friend who is holding such a hand? Are there any questions that will determine your future line of questioning? How would you reverse those questions and identify a constructive process?

Here's our constructive procedure. How does it compare to yours? Is it exactly the same? Is it equivalent somehow? Did we just partition the set of hands in a different order? Did we get the same final answer? Why or why not? Seriously, do *not* be discouraged if we differ in steps or final answer. It will be far more instructive for you to sit down and think about *why* our answers are different than to just read our correct solution. Seriously.

1. $Q\spadesuit$ present:
   (a) Choose the Queen of Spades: 1 option
   (b) From the remaining 51 cards that are not Queens (of which there are 3) and not Spades (of which there are 12 non-Queens), choose 4: $\binom{51-3-12}{4} = \binom{36}{4}$

   OR

2. $Q\spadesuit$ not present:
   (a) Choose a non-Spade Queen: $\binom{3}{1}$ options

(b) Choose a non-Queen Spade: $\binom{12}{1}$ options

(c) From the remaining 50 cards that are not Queens (of which there are 3) and not spades (of which there are 11 non-Queens and non-chosen), choose 3: $\binom{50-3-11}{3} = \binom{36}{3}$ options

Since the selections at each step yield unique outcomes, the Rule of Product applies, and since every hand with these properties either has $Q\spadesuit$ or doesn't, the Rule of Sum applies. Therefore, the number of hands with exactly one Queen and exactly one Spade is

$$\binom{36}{4} + \binom{3}{1}\binom{12}{1}\binom{36}{3} = 58,905 + 257,040 = 315,945$$

This is a trickier problem than the previous examples, so we encourage you to read over this proof multiple times until you are fully comfortable with it. In fact, ask your friend if he/she can solve the problem, and then try to convince them of your answer by following the steps of this proof. Do you understand them well enough to be able to explain them to someone else? If so, you are a master of combinatorial arguments!

In the next subsection, we seek to further develop your comfort with combinatorial arguments and proofs. Along the way, we will also introduce some standard combinatorial objects, so that we can count something other than poker hands. Counting questions about a deck of cards are common and easy to ask, but we'd like to talk about other stuff, too!

8.3.3 Other Counting Objects

n-Tuples from $[k]$

A deck of cards is a nice, standard, *physical* set of objects to count. Most people are familiar with them, and the fact that each card has *two* properties–suit and rank–allows for many interesting combinatorial problems to be posed. A more "abstract" example of a standard set of objects to count involves lists of natural numbers with specified lengths. We will make the following definition to allow us to refer to these sets in a concise form.

Definition 8.3.9. *Let $n, k \in \mathbb{N}$ be given. Then*

$$T_{k,n} = [k]^n = \{(a_1, a_2, \ldots, a_n) \mid \forall i.\ a_i \in [k]\}$$

That is, $T_{k,n}$ is the set of all n-tuples whose elements belong to $[k]$.

Note: We chose the letter T because these objects are *n-tuples*, i.e. ordered lists of length n. We will also point out that when k is a small number, like 2 or 3, it is common to replace the set $[k]$ with $[k-1] \cup \{0\}$. For instance, the concept of a *binary* n-tuple is quite common in mathematics, in part due to its prevalence in comptuer science. With that in mind, the case where $k = 2$ often considers ordered lists of length n whose elements are drawn from the set $\{0, 1\}$,

instead of $\{1, 2\}$. Since we are interested in *combinatorial* aspects of these sets (i.e. "How *many* sequences with property P are there?") we don't, in fact, care which convention is chosen. It is actually a simple exercise to prove that

$$|T_{k,n}| = |[k]^n| = k^n = |([k-1] \cup \{0\})^n|$$

by establishing a bijection between the underlying sets, $[k]$ and $[k-1] \cup \{0\}$. We will leave this for you to do ☺

Many counting arguments we will see in the next section can be conveniently phrased in this framework by identifying an appropriate k and n and an additional property that the ordered lists must have. For now, let's investigate a couple of simple cases and explore some applications. In each case, we will be looking at some subset, $S \subseteq T_{k,n}$, whose elements have a certain property (or properties); specifically, we will be looking to find $|S|$ by counting the elements of S. We will study some very simple cases first, then progress into some more challenging ones. The exercises in this section will explore these ideas even further.

Example 8.3.10. Let $n = 4$ and $k = 3$.

(1) What is $|T_{3,4}|$?

To count all the elements of $T_{3,4}$, we can construct this set via a four step process, where the i-th step corresponds to selecting the i-th element in the 4-tuple. At each such step, we have 3 options (each element is one of $\{1, 2, 3\}$), so the Rule of Product tells us there are $3 \cdot 3 \cdot 3 \cdot 3 = 3^4 = 81$ total elements of $T_{3,4}$. (Note: See the exercises, which ask for a proof that $|T_{n,k}| = n^k$, in general.)

(2) How many elements of $T_{3,4}$ have no 1s?

To count the elements of $T_{3,4}$ with no 1s, we can alter our 4 step process by restricting the number of options in each step. That is, each of the 4 positions of any element of $T_{3,4}$ with no 1 can only be filled from the set $\{2, 3\}$. Thus, the Rule of Product says there are $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$ such elements.

(3) How many have exactly one 1? Exactly two 1s? Exactly three 1s? Exactly four 1s?

To count the elements of $T_{3,4}$ with exactly one 1, can we use the same idea as the previous paragraph? Not exactly! The number of options available at each step in our process might *change*, depending on whether a 1 has already been placed in our 4-tuple. We must find a new approach. Instead, let's consider placing a 1 somewhere in our 4-tuple, then filling the remaining spots with elements from $\{2, 3\}$. That is, our four step process to construct a 4-tuple with the property that it has exactly one 1 is as follows:

(a) Choose one of the 4 spots to be occupied by the 1: $\binom{4}{1} = 4$ options. Then, for the remaining 3 unfilled spots, read left to right.

(b) For the first unfilled spot, select an element from $\{2, 3\}$: 2 options

(c) For the second unfilled spot, select an element from $\{2, 3\}$: 2 options

(d) For the third unfilled spot, select an element from $\{2, 3\}$: 2 options

Thus, there are $4 \cdot 2^3 = 32$ such elements of $T_{3,4}$.

Perhaps we were a bit verbose in this argument. We could have had two steps, where the first identifies where the 1 is placed and the second chooses from $\{2, 3\}$ for each of the remaining 3 spots. This is just a matter of semantics, though, and amounts to the same proof of the same fact. We presented these extra details to ensure that you follow our argument and understand the underlying principles. This will help you adapt these ideas to your own proofs!

We can use a similar argument to count the number of elements of $T_{3,4}$ with exactly 2 ones. The only difference is in Step 1: we must select 2 of the 4 spots to be filled with 1s. There are $\binom{4}{2}$ ways to do this. Then, there are two spots to be filled from $\{2, 3\}$. Thus, there are

$$\binom{4}{2} \cdot 2^2 = 24$$

such elements of $T_{3,4}$.

We will leave it to you to verify that there are 8 elements of $T_{3,4}$ with exactly three 1s, and 1 element with exactly four 1s. We will also leave it to you to verify and explain why it makes sense that $16 + 32 + 24 + 8 + 1 = 81$. (Challenge problem: can you generalize this result to any n and k?)

Example 8.3.11. Let $n \geq 3$. Count the number of binary n-tuples that have (a) exactly three 1s; (b) at least three 1s; (c) an even number of 1s.

Our context here is the set $\{0, 1\}^n$ of all n-tuples whose elements are drawn from the base set $\{0, 1\}$. (Notice that this is not *exactly* the set $T_{2,n}$ as defined above, but we explained how these sets are equivalent, in the sense that we can find a bijection between them.)

To answer question (a), we employ the same technique as the previous example. First, we select 3 of the n total spots to be filled with 1s; second, we fill the remaining $n-3$ spots with 0s. There are $\binom{n}{3}$ ways to complete the first step and, from there, the second step is deterministic (i.e. there is 1 way to do it), so there are $\binom{n}{3}$ binary n-tuples with exactly three 1s. (Note: we only specified $n \geq 3$ to ensure our answer is nonzero. If $1 \leq n \leq 2$, then we certainly can't have any such tuples! Indeed, this "verifies" that $\binom{n}{\ell} = 0$ whenever $\ell > n$.)

To answer question (b), we employ the same technique as in (a), but generalize from 3 to an arbitrary natural number ℓ. That is, we can count the number of binary n-tuples with exactly ℓ 1s, as follows: select ℓ of the n spots to be filled with 1s, then fill the remaining spots with 0s. To have at least three 1s, we must have either exactly three 1s, or exactly four 1s, or ..., or exactly n 1s. More rigorously, for every ℓ between 3 and n (inclusive), let A_ℓ be the set

of binary n-tuples with exactly ℓ 1s. Every binary n-tuple with at least three ones belongs to *exactly* one of the A_ℓ sets. Thus, we have identified a *partition* of the set of tuples we want to count, based on exactly how many 1s there are. Accordingly, the number we seek, by the Rule of Sum, is

$$\sum_{\ell=3}^{n} |A_\ell| = \sum_{\ell=3}^{n} \binom{n}{\ell}$$

You might be wondering whether this answer, in summation notation, is *acceptable*. In some sense, it is; not 10 minutes ago, we had no idea how many binary n-tuples there were with at least three 1s, and now we have a much better sense of that number. However, the solution, as presented, is more of a *method* for finding the precise number. If someone came up to you on the street and said, "Quick! Tell me the number of binary n-tuples with at least three 1s!", what would you do? You'd say, "Hold on, I just need to sum this series by evaluating each of the terms individually and then adding ..." Not ideal, right? It would be *nicer*, more convenient, to have a simple form of the solution; perhaps we could write it as just one binomial coefficient, or a sum/difference/product/quotient of two or three or some *small* number of such coefficients. That way, no matter what n is (i.e. no matter how large it becomes), we know we can always calculate the answer efficiently in a few short steps; moreover, we want to know that the *number* of such steps does *not* increase as n increases. With the summation form above, the number of terms in the sum grows as n does. This is not ideal.

We will leave the details for you to verify and explain, but we claim that an appropriate partition of the set of *all* binary k-tuples can be established to prove that

$$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \sum_{\ell=3}^{n} \binom{n}{\ell}$$

by invoking the Rule of Sum. (In fact, the proof of this equality delves into some of the techniques we will discuss in the next section, but we believe you can understand what the terms of this equation mean and why the equality must hold.) What we want to emphasize, though, is that we can rearrange that equation to obtain a *better*, in some *qualitative* sense, form of the original solution we sought:

$$\sum_{\ell=3}^{n} \binom{n}{\ell} = 2^n - \binom{n}{0} - \binom{n}{1} - \binom{n}{2}$$

Look at what we have achieved! No matter how large N is, we only have four terms to evaluate. The fact that this number is *fixed* is the main point. In fact, solutions of this form are given a name because we like them so much: **closed form**. The idea is that there is no "unnecessary" summation and the number of terms is fixed, regardless of the value of the variables contained therein.

In general, with combinatorics problems, we are always looking for a **closed form** of the solution, whenever possible. Sometimes, it is easy to come up with

a non-closed (some might say "open") form of a solution, but it might take some ingenuity to simplify this to a closed form. In this specific example, we relied on our observation that all binary n-tuples can be classified as having exactly zero or exactly one or exactly two or at least three 1s. This closed form of the solution not only allows us to evaluate the expression quicker and more easily, but it also provides some insight into even more of the underlying structure of the problem. For these reasons, we will always ask for closed form solutions.

Now, let's not forget about question (c)! To answer it, we employ a similar technique as in (b) and partition the set of binary n-tuples with an even number of 1s into those with exactly zero 1s (remember, 0 is an even number!), those with exactly two 1s, those with exactly four 1s, and so on. We have to be careful about the upper limit, though, because n is not necessarily even, itself! Recall the *floor function* that rounds a number down to the largest integer smaller than that number; an example is $\lfloor 5.7 \rfloor = 5$. With this in mind, we claim that there are

$$\sum_{\ell=0}^{\lfloor n/2 \rfloor} \binom{n}{2\ell}$$

binary n-tuples with an even number of 1s. We will leave it to you to fill in a proper explanation of this claim. Try presenting it to your friend and convincing them it is correct. Don't give up until they're fully convinced! We will forgo trying to find a *closed form* of this solution because it is not within our grasp ... or is it? See what you can deduce about this summation! What happens when n is even? When n is odd? Are there similar sums you can relate this expression to? What can you conclude?

Example 8.3.12. Count the number of binary 4-tuples whose 1s occur in pairs. To clarify, we would want to include $(1, 1, 0, 0)$ and $(1, 1, 1, 1)$ in our count, but *neither* $(1, 0, 1, 1)$ nor $(0, 1, 1, 1)$, for instance. Adapt this argument to count the number of binary 5-tuple whose 1s occur in pairs. Can you continue and find a general pattern?

To have 1s in pairs in a 4-tuple, this means we can either have 0 or 2 or 4 1s, in total. Let's define the sets S_0, S_2, and S_4 to be the sets of binary 4-tuples whose 1s occur in pairs *and* that have exactly 0 or 2 or 4 total 1s, respectively. (Note: we are not defining S_2 to be the set of binary strings with exactly two 1s, only. That would erroneously count a string like $(1, 0, 1, 0)$.) These sets form a partition of the set of elements we want to count, overall, so we merely need to count the elements of these three sets and add those numbers, by applying the Rule of Sum.

- To find $|S_0|$, we need to count the binary 4-tuples with no 1s, and there is only 1 such tuple: $(0, 0, 0, 0)$.

- To find $|S_2|$, we need to count the binary 4-tuples with two consecutive 1s and the remaining spots filled by 0s. Writing out the cases by hand,

$$(1, 1, 0, 0) \qquad (0, 1, 1, 0) \qquad (0, 0, 1, 1)$$

it is obvious that there are only 3 such tuples. (Can you identify a craftier argument for why there are 3? We'll come back to this when we look at 5-tuples.)

- To find $|S_4|$, we need to count the binary 4-tuples with four 1s. Certainly $(1, 1, 1, 1)$ is the only such tuple.

By the Rule of Sum, then, there are $1 + 3 + 1 = 5$ binary 4-tuples whose 1s occur in pairs.

To answer this same query about 5-tuples requires a little bit more ingenuity. We will define the same sets, S_0, S_2, S_4, amending the definitions to include 5-tuples (instead of 4-tuples). Still, this collection of 3 sets forms a partition of the set of binary 5-tuples we seek to count. It suffices to count each set's elements and apply the Rule of Sum.

- $|S_0| = 1$ because only 1 binary 5-tuple has no 1s: $(0, 0, 0, 0, 0)$.

- To find $|S_2|$, we can again write out the cases by hand, but it would be nice to come up with an argument we can easily adapt to k-tuples, for any $k \in \mathbb{N}$. To have one pair of 1s and the remaining spots filled by 0s, we can think of the block "1, 1" as a single unit, placed amongst three 0s. Thus, we are really counting how many ways we can place a single, special unit into an ordered list of length 4 and then deterministically fill the remaining spots with another fixed element. Certainly, there are 4 ways to do this, and writing out the cases by hand verifies this claim. (What we are really doing is noting that a consecutive pair of 1s is determined by the index in the tuple of the *first* 1 of the pair; that index can be any of $\{1, 2, 3, 4\}$, so there are $\binom{4}{1} = 4$ options.)

- This technique applies when we consider $|S_4|$, as well. Here, we have two separate, consecutive pairs of 1s, so we can treat each pair as a single block, "1, 1". Thus, we really have two "1, 1" blocks and one 0 to place in an ordered list of length 3. Since the 1, 1 blocks are identical, we can count these ordered lists by selecting the two spots for the 1, 1 blocks. Therefore, there are $\binom{3}{2} = 3$ such tuples. (Equivalently, we can think of selecting the spot for the 0, i.e. $\binom{3}{1} = 3$ options.)

Therefore, there are $1 + 4 + 3 = 8$ binary 5-tuples whose 1s occur in pairs.

Alphabets and Words

Related to the idea of k-tuples with elements drawn from $[n]$ is the idea of creating *words* from a given *alphabet*. There is not much of a difference between these two concepts (and mathematically, they are really *equivalent*, so there's nothing new!) but it allows for different terminology and makes pertinent connections to some "real-world" concepts and problems. For this reason, and the fact that you might find it easier to work with one formulation or the other, we present this subsection and make connections to the previous subsection.

We will introduce and motivate the ideas of this subsection with some examples. In each example, we will specify an *alphabet*, whose elements are the allowable *letters* we can use to construct *words*. By "word", though, we really mean *any* ordered string of letters drawn from the given alphabet. So, for instance, we use the standard English alphabet in the first example; in that case, ZPQ is a perfectly acceptable three-letter word (but good luck trying to pronounce it!). The main reason we allow for this generality is to avoid any semantic or etymological arguments, like whether or not REALISE should be an acceptable variant of REALIZE or if ZZZ really belongs as an onomatopoetic interjection. This means we have to strip away some of the connotations around the word "word" and think of it as just a string of letters with no other inherent meaning besides its components and their order.

Example 8.3.13. Let's consider the standard, 26-letter English alphabet in this example.

1. How many 3-letter words can be made this alphabet?

   Think about how this is *identical* to asking about $T_{26,3}$. We have 26 allowable letters and want to form ordered lists of length 3. By establishing a bijection between the sets $\{A, B, C, \ldots, Z\}$ and $[26]$ (which is actually a common and simple substitution cipher you might have played with as a child), we can rigorously show how this question is equivalent to asking what $|T_{26,3}|$ is.

   Without necessarily making this connection, though, we can easily count the 3-letter words by noting that constructing one amounts to a 3 step process (fill in 3 letters, left to right) with 26 options at each step. Thus, there are 26^3 three-letter words.

2. How many 4-letter words can be made?

   By the same logic as the previous example, there are 26^4 four-letter words.

3. How many n-letter words?

   We'll let you handle this one ☺

4. How many 4-letter words start with a vowel?

   Note: we consider {A,E,I,O,U} to be the set of vowels (i.e. Y is never included, despite what the alphabet song might tell you). With that in mind, we can amend the four-step process of constructing a four-letter word to guarantee a vowel occurs in the first position on the left. There are 5 ways to complete that step, followed by 26 ways for each of the next 3 steps. Thus, there are $5 \cdot 26^3$ four-letter words that start with a vowel.

5. How many 4-letter words have at most 2 consonants?

   A consonant is a non-vowel, so there are $26 - 5 = 21$ consonants in the alphabet, under our definitions. To have at most two consonants means we have exactly 0 or exactly 1 or exactly 2, so we can partition the set

of words in question into three corresponding sets, S_0, S_1, S_2, and count each separately. By applying the Rule of Product, we find that

$$|S_0| = 5^4$$
$$|S_1| = \binom{4}{1} \cdot 21 \cdot 5^3$$
$$|S_2| = \binom{4}{2} \cdot 21^2 \cdot 5^2$$

and, therefore, there are

$$5^4 + 4 \cdot 21 \cdot 5^3 + \binom{4}{2} \cdot 21^2 \cdot 5^2$$

four-letter words with at most two consonants. (Challenge problem: How many four-letter words have at least three consonants? Use this to make a claim about the number 26^4.)

What is wrong with the following argument about S_2?: Construct a four-letter word with exactly two consonants by selecting two consonants, then selecting a position for the first consonant, then selecting a position for the second consonant, then filling the remaining two spots with vowels, yielding

$$\binom{21}{2} \cdot 4 \cdot 3 \cdot 5^2$$

such words.

Think carefully about this. Remember, these "find the flaw" questions are not merely looking for you to identify that there *is* an error, but also to explain *why* it is an error and how it could be fixed.

6. How many 4-letter words have 4 distinct letters?

   Without any pre-thought, we can jump into answering this one by describing a four-step process of filling the four letters left to right, and reducing the number of options by one at each step. Thus, there are

   $$26 \cdot 25 \cdot 24 \cdot 23$$

   four-letter words with four distinct letters.

   Does this look familiar, though? Look back to where we defined *arrangements*. This is precisely the idea we are using here! From a set of 26 elements, we want to construct an ordered list of length 4 with no repetition, i.e. a 4-arrangement from a set of 26 elements. The formula we derived tells us there are exactly

   $$\binom{26}{4} \cdot 4! = \frac{26!}{4!\ 22!} 4! = \frac{26!}{22!} = 26 \cdot 25 \cdot 24 \cdot 23$$

such arrangements. That's the lesson of this example: by taking advantage of previously defined terms and derived formulas, and relating a current question to those ideas, we can "jump" to a solution.

7. How many 4-letter words have exactly one letter repeated twice?

   To construct a word with these properties, we need to know which letter is repeated, and where its two instances occur, as well as the other two letters that appear in the word. Thus, we identify a three-step process: (1) select the repeated letter; (2) select 2 of the 4 open spots for that letter; (3) *arrange* two of the remaining 25 letters in the remaining two spots:

   $$\binom{26}{1}\binom{4}{2}\binom{25}{2}\cdot 2!$$

8. How many 5-letter words have exactly two letters repeated twice each?

   Following similar logic as the previous example, we can (1) select two letters to be repeated; (2) select two of the 5 open spots for the first (alphabetically speaking) repeated letter; (3) select two of the remaining 3 spots for the second (alphabetically) repeated letter; and (4) select one of the remaining 24 letters to fill the final, fifth spot. Thus, there are

   $$\binom{26}{2}\binom{5}{2}\binom{3}{2}\binom{24}{1}$$

Example 8.3.14. How many rearrangements (i.e. permuations) of the alphabet put U and ME together? ☺

You might think to approach this problem with a "subtraction" idea; that is, you might try to count all of the rearrangements of the alphabet that put the letter U *not* next to the letters ME (in that order). You should try working this out and see where it takes you. What we will do, though, is present a *different* and, we believe, shorter solution. The idea behind this solution will be useful in other problems, and it boils down to treating the two-letter word ME as a single *block*, just like any other single letter.

To reverse the question a little bit, instead of asking how many words there are of some type from a given alphabet, we might wonder how many rearrangements (i.e. *anagrams*) there are of a given word. Care must be taken in considering these questions because when letters are repeated, things can get tricky! For instance, how many anagrams are there of the word A? How about the word AAAAA? How about AABBBCCCCDD? Exactly!

Example 8.3.15. Let's start with a simple case. How many anagrams of the word HEART are there? Remember that we count *all* permutations of the letters as an acceptable word, so leave behind your Scrabble thought processes ☺(By the by, the Scrabble answer to this question is 4–HEART, HATER, EARTH, RATHE.) Since each letter is *distinct* in this word, the answer is simple: we merely count all of the permutations of the 5 letters. Accordingly, there are $5! = 120$ anagrams of HEART.

Now, how many anagrams of APPLE are there? Notice that the letter P appears twice, so we can't really consider permutations of a *five*-element set. If we were to do that, each word would actually occur *twice*; that is, APPLE and APPLE would both be counted. Do you see the difference between those two words? We just switched the Ps! Of course, these are the *same* word, so we have to factor this in to our consideration of permutations.

How do we do this? One helpful trick is to "label" the repeated Ps. Reading left to right in APPLE, let's call the first one P_1 and the second one P_2. This will help us to sort out the repeated permutations. We now have five distinct elements in our word–A, P_1,P_2,L,E–so we can consider all 5! permutations of these elements. We know that this *overcounts*, though, so we would like to figure out how *by how much* this overcounts. Define G to be the set of anagrams of APPLE (so we're trying to identify $|G|$) and let M be the set of permutations of the five distinct elements listed above. How are $|G|$ and $|M|$ related?

To answer this question, we can think of constructing the elements of M from elements of G. Specifically, we can construct any element of M (a permutation of the 5 distinct elements) by first taking an element of G (an anagram of APPLE) and labeling the Ps. However, this won't generate *all* elements of M. To do this, we have to take elements of G and, from them, construct two elements; specifically, we must label the Ps and then consider *both* of the orderings of the Ps within the word. Let's see an example:

- Take an element of G, say PAPEL.
- Label the Ps from left to right: P_1AP_2EL
- Construct both orderings of the Ps and count both of those words as elements of M: $P_1AP_2EL \in M$ and $P_2AP_1EL \in M$

Since there are $2! = 2$ ways to permute the two Ps within the word, we have shown that $|G| \cdot 2! = |M|$. We described a two-step process to generate elements of M and applied the Rule of Product. Accordingly, we can rearrange this equation to find the quantity we were looking for:

$$|G| \cdot 2! = |M| \implies |G| = \frac{|M|}{2!} = \frac{5!}{2!} = 60$$

For a slightly harder example, let's count the anagrams of COMBINATORICS. We want to apply the same strategy as the previous example and label the repeated letters to relate the anagrams to permutations of, in this case, 13 distinct elements. Again, let's define G to be the set of anagrams of COMBINATORICS and M to be the set of permutations of the 13 distinct elements in $\{C_1O_1MBI_1NATO_2RI_2C_2S\}$. We can describe a four-step process that generates elements of M:

1. Take an element of G and label the repeated letters, reading left to right: $|G|$ options
2. Permute the two repeated Cs: 2! options.

3. Permute the two repeated Os: 2! options.

4. Permute the two repeated Is: 2! options.

Thus, by the Rule of Product, we may conclude

$$|M| = |G| \cdot 2! \cdot 2! \cdot 2! \implies |G| = \frac{|M|}{2! \cdot 2! \cdot 2!} = \frac{13!}{2! \cdot 2! \cdot 2!}$$

You might wonder why we chose to write $2! \cdot 2! \cdot 2!$ instead of just 8. We find it more enlightening and illustrative to leave our answer in terms of factorials because it indicates where those terms *came from*, too.

What happens if a letter is repeated more than twice? The only difference occurs when we consider permuting the repeated instances of that letter. We will let you fill in the details of the argument, but we claim that the word AABBBCCCCDD has

$$\frac{11!}{2! \cdot 2! \cdot 3! \cdot 4!}$$

anagrams. Can you "see" why without going through the details completely yet? Can you fill in those details to confirm your intuition? Can you prove this fact and convince a friend? Try it!

There are several more anagram questions in the exercises. Later on, we will even prove a result that will generalize this technique of labeling repeated letters and accounting for their permutations.

Before moving on, we should point out an observation that is similar to something we've seen before. In some examples so far, we ended up *subtracting* a count from a total, and we pointed out that a sophisticated proof-writer would just state the subtraction idea directly, although we ask you to phrase it in terms of a *partition* and apply the Rule of Sum. Similarly, with the previous example we just did, we ended up *dividing* a count to eliminate some overcounting, but we made sure to phrase this in terms of a *process* and apply the Rule of Product; afterwards, we could algebraically divide to simplify. A sophisticated proof-writer would write probably write the same solution by taking the overcount and arguing that we can "divide out" to eliminiate the overcounting. This is dangerous, we say, and we require you to *not* do this (for now, in our contexts). If you allow yourself to make these kinds of arguments, it's too easy to erroneously "divide out" in a situation where it is unwarranted and incorrect to do so! Forcing yourself to work out these arguments "from the ground up" will more firmly solidify these underlying principles and allow you to more confidently apply "subtraction" and "division" principles later on in your mathematical careers. Just remember that we ask you to use only ROS and ROP in our contexts!

Let's look at two quick examples of alphabets and words that aren't based on the standard English alphabet, per se.

Example 8.3.16. A phone number in the United States consists of an area code (3 digits) and a local number (7 digits). The digits are drawn from the set

$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, but neither the area code nor local number can begin with a 0. How many possible phone numbers are there?

This is easy to count by establishing a 10 step process, corresponding to the 10 total digits of a phone number. Eight of the digits have 10 options, and two of them have 9 options (no 0), so the Rule of Product tells us there are

$$10^8 \cdot 9^2 = 8,100,000,000$$

possible phone numbers. This number is slightly larger than the current world population, so it seems like our system is safe, for now!

Example 8.3.17. Suppose a restaurant has a menu with three different categories: appetizers, entrees, and desserts. There are 5 appetizers, 9 entrees, and 4 desserts. You take your date to this restaurant and you pass the time waiting for your server t)o appear by figuring out how many possible orders you can make, imposing certain conditions. (Unfortunately, you forget to decide what to order and you end up choosing an order randomly, but that's beside the point.)

(1) How many possible orders can you make, assuming you must order one appetizer, one entree, and one dessert?

This is a simple application of ROP. There are $5 \cdot 9 \cdot 4 = 180$ possible orders. How is this like alphabets and words? Well, you can think of it as constructing a three-letter word, but the alphabet for each "slot" in the word is different.

(2) How many possible orders can the two of you make, assuming you each order one app, one tree, and one zert, but you two can't order the *same* thing in any category?

Think of this as a 3-step process, with each step divided into 2 parts. First, you order an appetizer, and then your date does, too (making sure to pick from amongst the appetizers you *didn't* pick). Second, you order an entree, and then your date picks a different one. Third, you order a dessert, and then your date picks a different one. Clearly, then, there are

$$(5 \cdot 4) \cdot (9 \cdot 8) \cdot (4 \cdot 3) = 20 \cdot 72 \cdot 12 = 17280$$

possibilities. Compare this to the number of possibilities *without* the restriction of ordering different items in each category, which we can find by using our work on question (1):

$$180 \cdot 180 = 32400$$

Again, we can think of this as a restrictive alphabet/words problem.

Balls and Bins

This is a common formulation of combinatorics problems in more advanced courses. It's particularly helpful when discussing *probability* and using combinatorics facts and ideas to explore probability. We would like to bring it up here

because it introduces the important distinctions between *distinguishable* and *indistinguishable* objects. To motivate this discussion, let's pose a seemingly simple question:

> Consider a bin containing n balls; how many ways can we select k balls?

What's your answer? If you said "$\binom{n}{k}$", you could be right. If you said "1", you could also be right. How is that possible?! Well, we didn't specify whether the n balls in the bin are *distinguishable*; that is, we didn't say whether or not they're all different, whether we can tell any two balls apart.

Imagine a bucket with 100 tennis balls in it. If we pulled out two balls and showed them to you, could you tell them apart? Maybe they have a different number of fuzzy yellow hairs on them, or maybe they're different brand names, or something like that ... but maybe we can't do that. Maybe all of the balls are completely identical. In that case, it doesn't matter "which" k balls we pull out, because we can't tell them apart. All "possible" selections of k balls amount to the same thing, so the answer of "1" makes total sense. However, if all of the balls had a distinct number on them, or they were all different colors, or ... imagine any distinguishing property you'd like. In any of those cases, "$\binom{n}{k}$" is the correct answer. For these reasons, the originally posed question was a poor one; we should have been specific about whether the balls are *distinguishable* or not.

This idea of *distinguishability* has come up before. Remember that grid of counting formulas we established back in Section 8.2.3? One of the essential questions in the grid was whether the order of a selection/arrangement *distinguished* the outcome. For example, a selection does not care about order. The selections $\{1, 3, 4\}$ and $\{3, 4, 1\}$ are identical (you should also think of them as *sets* to understand this) because the order in which the elements are written does *not* distinguish them. Conversely, the *arrangements* $(1, 3, 4)$ and $(3, 4, 1)$ are different because the order of the elements *does* distinguish them.

In the context of a "balls and bins" problem, we will specify the distinguishability of items by saying they are numbered or colored somehow. This might also involve a mix of distinguishable/indistinguishable features, though, so be careful! The next example illustrates this interplay.

Example 8.3.18. Suppose we have a bin containing balls that are colored red, blue, or green (i.e. each ball has one color of those three). There are 3 balls of each color in the bin, and any two balls of the *same* color are indistinguishable. We pull out four balls. How many possible outcomes are there?

Try playing around with this problem before reading our solution. You might come up with your own method of solving it!

A first approach here might be to just enumerate the possibilities and then

try to infer a pattern. We might strart writing out the outcomes as:

3 Red and 1 Blue
3 Red and 1 Green
2 Red and 2 Blue
2 Red and 2 Green
2 Red and 1 Blue and 1 Green
$\vdots$

and so on. Notice how we are keeping track of this information, though; every outcome is characterized by (a) how many *red* balls we picked, (b) how many blue balls, and (c) how many green balls. In essence, we can characterize every outcome by an ordered 3-tuple of the form (r, b, g) where r is the number of red balls, and similarly for b and g. The only condition we require is that $r+b+g = 4$ and each value satisfies $0 \leq r \leq 3$, $0 \leq b \leq 3$, $0 \leq g \leq 3$. We really just need to count how many 3-tuples satisfy those conditions! We can split this count into how many nonzero terms there, and analyze each case separately.

- If two terms are 0, then the third term must be 4. There are $\binom{3}{1} = 3$ choices for which term is nonzero, so there are 3 such possibilities.

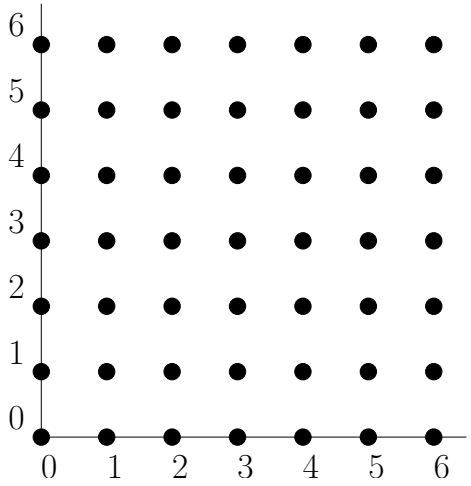
- If one term is 0, then the other two terms must sum to 4 and both be nonzero. There are 3 ways to do this: $1 + 3$ and $2 + 2$ and $3 + 1$. Since there are 3 choices for which term is 0, by ROP, there are $3 \cdot 3 = 9$ such possibilities.

- If all three terms are nonzero, then we see that the only such sum is $1+1+2$, in some order. There are 3 choices for which term is 2, and then the other terms must be 1. Thus, there are 3 such possibilities.

By ROS, there are $3 + 9 + 3 = 15$ total possibilities.

Lattice Paths

Consider the set $\mathbb{N} \cup \{0\} \times \mathbb{N} \cup \{0\} = (\mathbb{N} \cup \{0\})^2$ consisting of all ordered pairs of natural numbers or 0.In fact, let's represent this set visually on the plane:



This "grid" of dots on the plane is known as a *lattice*. Here's a natural question: Given any point in the lattice, how many ways are there to "get there" from the origin, $(0,0)$? Let's be more specific. Let's define a **lattice path** to be a path from $(0,0)$ to a particular point that is only allowed to move *rightwards* or *upwards* at any step. This is what the next definition conveys:

Definition 8.3.19. *Let $(x,y) \in (\mathbb{N} \cup \{0\})^2$. A* **lattice path** *to (x,y) is an ordered tuple of points in the plane lattice where the first element of the tuple is $(0,0)$, the last element of the tuple is (a,b), and every element in the tuple only differs from the previous one by having exactly one coordinate that is exactly one larger than the corresponding coordinate of the previous element.*

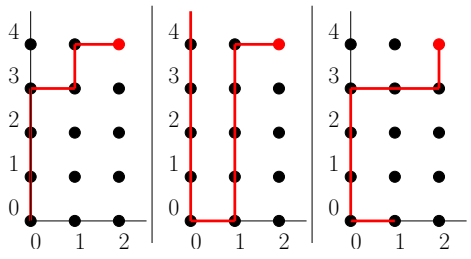
More rigorously, given (x,y) a lattice path is an n-tuple $(P_1, P_2, \ldots, P_n)$, for some $n \in \mathbb{N}$, where each $P_i = (x_i, y_i)$ is a point in the lattice, and

$$\forall i \in [n-1] \bullet (x_{i+1}, y_{i+1}) = (x_i + 1, y_i) \ \vee \ (x_{i+1}, y_{i+1}) = (x_i, y_i + 1)$$

and, furthermore, $(x_1, y_1) = (0,0)$ and $(x_n, y_n) = (x,y)$.

That is, a lattice path is a sequence of points in the lattice from $(0,0)$ to (n,n) where we are only allowed to move rightwards *or* upwards *by one grid point at every step.*

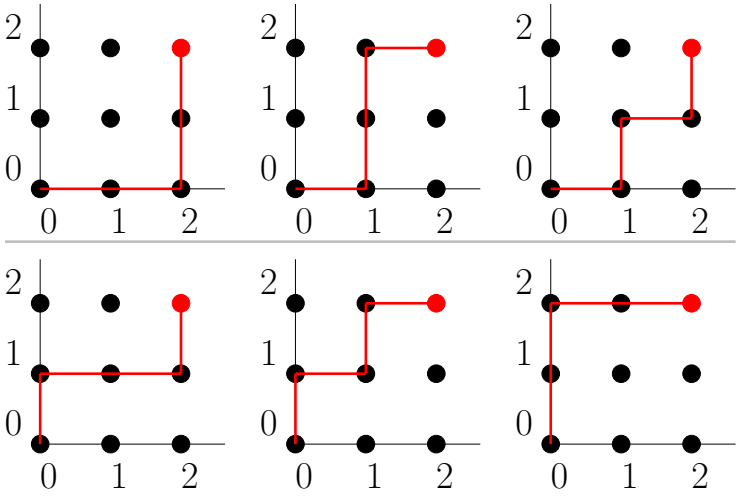
Example 8.3.20. Consider the point $(2,4)$ in the plane lattice. In the diagram below, we plot a few sample lattice paths to $(2,4)$.



Our question is as follows:

> Given $(a, b) \in (\mathbb{N} \cup \{0\})^2$, how *many* distinct lattice paths are there to (a, b)?

To begin to answer this, let's look at a simple case with small values so we can actually enumerate all of the paths. Let's consider lattice paths to $(2, 2)$:



How might we represent lattice paths in a *combinatorial* way? That is, how can we represent them in a way that allows us to conveniently count some objects? Think about the defining aspects of a lattice path: every "move" in the construction of a lattice path must be *rightwards* or *upwards*. It would make sense, then, to somehow represent when we make a "Right" move and when we make an "Up" move. Then, we just need to count how many sequences of choices of "Right" and "Up" actually bring us to the point (x, y) in question.

That's easy, though! What characterizes the point (x, y) on the plane? Well, it's x grid points to the right of $(0, 0)$ and y grid points up from $(0, 0)$. Thus, no matter what our path *looks like*, we know there must be x rightward moves and y upward moves. Look back at the 6 lattice paths to $(2, 2)$ above. Think about following the path, starting at $(0, 0)$, and writing down R or U at each grid point, depending on where we go next. This yields the following 6 sequences of Rs and Us

$$RRUU\,,\ RUUR\,,\ RURU\,,\ URRU\,,\ URUR\,,\ UURR$$

What properties do these sequences share? Each one has 2 Rs and 2 Us, since we must end at $(2, 2)$, and so each sequence has 4 terms, in total. Notice that this is much like a restricted alphabet/words problem: we want to find the number of words of length 4, drawn from the alphabet {R,U}, that contain exactly two of each letter!

In general, then, we know that any lattice path to (x, y) can be represented by a $(x + y)$-tuple of Rs and Us with exactly x Rs and y Us. To identify how *many* such sequences there are, we have a two step process:

1. From $x + y$ empty slots, choose x of them to be filled with Rs: $\binom{x+y}{x}$ options
2. Fill the remaining $(x + y) - x = y$ slots with Us: 1 option (deterministic)

Thus, we have the following result.

Proposition 8.3.21. *For every $(x, y) \in (\mathbb{N} \cup \{0\})^2$, there are exactly $\binom{x+y}{x}$ lattice paths from $(0, 0)$ to (x, y).*

We will explore some interesting applications and properties of lattice paths in the exercises. For now, we want to point out their existence and their relationship with sequences and selections. But here's one more observation about them: Why did we choose to count the number of sequences of length $x+y$ with exactly x Rs? Would it be any different to count the the number of sequences of length $x + y$ with exactly y Us? Think about it: every lattice path to (x, y) needs exactly x Rs and exactly y Us, and ensuring one of those properties holds guarantees the other will, as well. Thus, we could have presented the following result:

Proposition 8.3.22. *For every $(x, y) \in (\mathbb{N} \cup \{0\})^2$, there are exactly $\binom{x+y}{y}$ lattice paths from $(0, 0)$ to (x, y).*

This not only *proves* the following fact

$$\binom{x+y}{x} = \binom{x+y}{y}$$

but it also introduces us to a new and helpful proof strategy: **Counting in Two Ways**. We identified one *set* of objects (the set of lattice paths from $(0, 0)$ to (x, y)) and proceeded to explain two *different* ways of counting that same set of objects. Each way yielded a different expression for the cardinality of that set, and we can therefore declare those two expressions to be equal. This first example illustrates the main idea behind Counting in Two Ways, and we will explore several more examples, and the general technique, in the following section.

8.3.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) How can you identify that a proposed counting argument is an **undercount**? How can you show it is an **overcount**?

(2) Explain the relationship between "$k-$tuples from $[n]$" and "Alphabets and Words". How are they fundamentally the same?

(3) Say we are selecting k balls from a bin with n balls. Why does it matter whether the balls are *distinguishable*?

(4) Why is it that the number of lattice paths from $(0,0)$ to (x,y) is equal to *both* $\binom{x+y}{x}$ and $\binom{x+y}{y}$?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Find the number of 5-card poker hands that are **two pair**, and prove your claim.

(2) Find the number of 5-card poker hands that are a **full house**, and prove your claim.

(3) How many anagrams of the word COMBINATORICS are there? What about of MASSACHUSETTS?

(4) Consider finding the number of 4-tuples from $\{1,2,3\}$ that include at least one of each number. For each of the following proposed "proofs", show that it is **incorrect** by exhibiting such an object that has been counted **twice** by the proposed argument.

   (a) Pick one of the 4 spots in the tuple for the 1, then pick a spot for the 2, then pick a spot for the 3. Then, pick one of the three elements to appear in the 4th empty spot.

$$\binom{4}{1}\binom{3}{1}\binom{2}{1}\binom{3}{1} = 72$$

   (b) Pick 3 of the 4 spots to be filled with the elements 1,2,3. Permute those elements in those chosen spots. Pick a number for the 4th empty spot.

$$\binom{4}{3}3! \cdot 3 = 72$$

(5) For this problem, consider a *word* to be any string of English letters, whether or not it actually spells something in the dictionary. For instance, *ZYQFIB* is a valid *word* of length 6.

(a) How many words of length 2 are there?

(Answer this question in *two* ways: with an exponential number, and with a sum of two terms.)

(b) How many words of length 7 have exactly 3 As?

(c) How many words of length 7 have at most 2 vowels? (Note: A, E, I, O, U are vowels. Y is not.)

Consider the set S_n of all binary strings of length n. For each of the following stated properties, count how many elements of S_n have that property. (Note: Each property is separate; don't consider satisfying all of them at once, say.)

(a) Exactly 3 positions are 0s.

(b) At most 3 positions are 0s.

(c) At least 4 positions are 0s.

Note: Use the last two parts to write 2^n as a sum of binomial coefficients!

(d) More positions are 0s than are 1s.

(6) Let $n \in \mathbb{N}$ be given. How many lattice paths go from $(0, 0)$ to $(2n, 2n)$? How many such paths *also* go through (n, n)?

(7) Consider the following explanation:

> The number of 6-card hands, as dealt from a standard deck of cards, that have *at least one* card from each of the four suits is
>
> $$\binom{13}{1}\binom{13}{1}\binom{13}{1}\binom{13}{1}\binom{48}{2}$$
>
> because we select one card from each of the four suits and then, from the remaining 48 unused cards, select two more.

Is this count correct? If you think it is an *overcount*, exhibit a specific hand and show how it is counted in two ways. If you think it is an *undercount*, exhibit a specific hand and show how it is not counted.

8.4 Counting in Two Ways

If you're just jumping into this section, reread the last example from the previous section because it provides a perfect introduction to (and example of) Counting in Two Ways. In that example, we counted the number of lattice paths to a particular point in *two* different ways, deducing that the two expressions we found must be equal. Specifically, we deduced that $\binom{x+y}{x} = \binom{x+y}{y}$. With that example already under our belt, we will outline a general strategy here and apply it to several examples. Along the way, we will not only practice this technique, but we will also be proving some useful combinatorial results that we can apply to other problems!

Let's start by actually presenting an *alternative* proof of the example from the previous section. There is a much shorter argument that doesn't delve into lattice paths at all and is a more memorable and understandable explanation of this result.

Proposition 8.4.1. *Let $n, k \in \mathbb{N} \cup \{0\}$. Then* $\binom{n}{k} = \binom{n}{n-k}$.

Proof. Let S be the set of subsets of $[n]$ with size k, i.e.

$$S = \{T \subseteq [n] \mid |T| = k\}$$

By the definition of k-selections, $|S| = \binom{n}{k}$, since constructing a set $T \subseteq [n]$ with $|T| = k$ amounts to selecting k elements from a set of n elements.

Equivalently, we can construct a set $T \subseteq [n]$ with $|T| = k$ by selecting $n - k$ elements to *not* include in T; this means $n - (n - k) = k$ elements have been selected to belong to T. The number of ways to do this is $\binom{n}{n-k}$. Since every such set T can be constructed this way, we have shown $|S| = \binom{n}{n-k}$.

Therefore, $\binom{n}{k} = \binom{n}{n-k}$. □

We find this to be a more memorable proof of this fact because we can summarize the entire proof in just one sentence rather nicely

> "Count the k-element subsets of $[n]$ by identifying the elements to include *or* the elements to omit."

This is the idea we remember; from it, we can reconstruct the proof. It doesn't make sense to try to "memorize" a proof sentence by sentence; rather, it is helpful to remember the *kernel* of the proof's main idea and then fill in the details.

8.4.1 Method Summary

Why It Works

Let's abstract one level and discuss Counting in Two Ways as a proof technique. Let's talk about *why* it works and *how* to employ it. Then, we'll go through

several more examples. We touched on the *why* idea at the end of the previous section, so we will reiterate those ideas here. "Counting in Two Ways" is the best name for this proof technique because it explains the strategy in its own name! Any proof following this technique identifies a finite set of elements and offers *two ways* to count those elements. By using the Rules of Sum and Product, and other combinatorial results we have seen, those two ways yield different algebraic expressions for the same number, namely the *cardinality* of that set of elements in question.

A good proof clearly identifies the finite set to be counted and the two distinct ways of counting its elements, and then concludes by equating the two algebraic expressions. Thus, any result proved by this method will be some kind of *identity* or *equation* involving binomial coefficients, summations, and other algebraic expressions. The point of the proof is to explain those expressions clearly in terms of a counting argument, as opposed to a strict algebraic simplification of the terms.

Look at the result we just proved above: yes, we could directly verify that $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$, but where is the fun in that? That would not be an *interesting* proof, by any stretch of the word's meaning, nor does it provide any *insight* into why the result is true. Furthermore, as we investigate more and more challenging problems of this type, algebraic verification becomes rather difficult, and in some cases, pretty much impossible!

How To Use It

We will go on to present several examples (and non-examples) later in this section, but we want to present an outline of the Counting in Two Ways method first. This will provide us with a standard by which to measure future proofs of this style; we can read through them and make sure they follow the important points of structure and clarity and correctness. We will hold ourselves to these standards and expect you to do the same. We also present you with some standard combinatorial objects that are used in Counting in Two Ways proofs, and as we proceed with the examples, we will point out when to consider using particular sets of objects to count.

Now, here is a skeleton structure of *every good Counting in Two Ways proof*!

1. State the result to be proven.
   (Note: remember to quantify any variables that appear in the expression!)
2. Define a set—let's call it S—of objects to be counted.
3. Count the elements of S in one way by following a proper combinatorial argument. Equate the derived expression with $|S|$.
4. Count the elements of S in another way by following a proper combinatorial argument. Equate the derived expression with $|S|$.
5. Conclude that since both derived expressions equal $|S|$, they must be equal, as well.

That's it! Like we said, the technique's name is the technique itself, so it should be easy to remember. However, after reading many of these proofs over the years, we have noticed that certain mistakes are commonly made. We've listed the most common here. Think about why doing any of these would amount to a "bad" proof in some sense. Which property of a good proof does each mistake fail to satisfy? Correctness? Clarity? Brevity?

Common mistakes to avoid:

- Forgetting to define a set of objects to be counted.
- Defining a set, but counting something else in two ways.
- Counting one set of objects, but then counting a *different* set of objects in another way.
- Failing to equate the two expressions in a conclusion

Other mistakes might arise in the actual combinatorial proofs, rather than the technique, as a whole. But be on the lookout for those, too!

8.4.2 Examples

Let's go through several examples in full detail. This will help you get a handle on how the Counting in Two Ways technique is applied, provide you with some canonical examples to look back on and re-read, and also provide you with some fundamental combinatorial results that can be applied in future problems. In each example, we attempt to not only prove the result in question, but also explain *how* we come up with the proof, what our thought processes might be in constructing the argument, and how you might try to approach problems like these on your own. One of the beautiful aspects of counting in two ways proofs is that, by the end of such a proof, we can usually neatly summarize the proof's main idea. We will do that for each of the proofs we present, and encourage you to attempt the same summary at the end of any such proof you write. This makes the proof idea easier to remember, and will allow you to reconstruct the entire proof from just one sentence.

Proposition 8.4.2 (Pascal's Identity)**.** *For any $n, k \in \mathbb{N}$,*

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

Proof strategy: Seeing binomial coefficients, like $\binom{n}{k}$, indicates that we might want to count subsets of $[n]$ with particular sizes. The left-hand side of this identity is easy to represent (count all the subsets with size k), but what about the right-hand side? Seeing a *sum* of two terms indicates a partition of some kind. We must identify a certain property of subsets of $[n]$ with size k so that some subsets *do* have the property, and some of them *don't*. Noticing that the

only difference in the terms is in the "bottom coefficient", we can figure out a fruitful partition ... See if you can come up with it on your own before reading on!

Proof. Let $S = \{T \subseteq [n] \mid |T| = k\}$. By the definition of k-selections, we know $|S| = \binom{n}{k}$. Next, define the sets

$$A = \{T \subseteq [n] \mid |T| = k \wedge 1 \in T\}$$
$$B = \{T \subseteq [n] \mid |T| = k \wedge 1 \notin T\}$$

Certainly, $A \cap B = \varnothing$ since both $1 \in T$ and $1 \notin T$ cannot be true for any set T. Also, $S = A \cup B$, since either $1 \in T$ or $1 \notin T$ is true for any set T. Thus, $\{A, B\}$ is a partition of S, and we know, then, that $|S| = |A| + |B|$.

To find $|A|$, we identify a two-step process for constructing elements $T \in A$: (1) include the element 1 in T; (2) from the remaining $n - 1$ elements, select $k - 1$ more to make a set of k elements. By the Rule of Product, we conclude

$$|A| = 1 \cdot \binom{n-1}{k-1} = \binom{n-1}{k-1}$$

Simiarly, to find $|B|$, we identify a two-step process for constructing elements $T \in B$: (1) omit the element 1 from T; (2) from the remaining $n - 1$ elements, select k elements. By the Rule of Product, we conclude

$$|B| = \binom{n-1}{k}$$

By the Rule of Sum, then, we conclude

$$|S| = |A| + |B| = \binom{n-1}{k-1} + \binom{n-1}{k}$$

By equating the two expressions for $|S|$, we conclude that

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

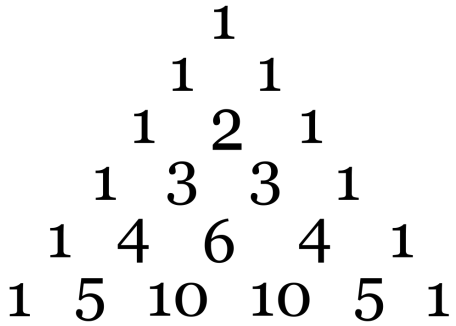
□

Proof summary: Count the k-element subsets of $[n]$ by partitioning based on whether or not a particular element (say, 1) belongs to the subset.

Question: What if we had used the element n, instead of 1, to construct our partition? Would the proof be any *different*, structurally? No! All that matters is we identified a *specific* elements, and defined the partition sets, A and B, based on that element.

Historical note: This Proposition is named for the French mathematician Blaise Pascal. Perhaps you have heard of Pascal's Triangle, as well? This triangle of natural numbers is constructed by writing two rows of 1s, all 1s along the

perimeter, and then filling in every other number with the sum of the two above it. Based on this Proposition we just proved does that give you any indication as to what the numbers contained in the triangle actually are? That's right, they're binomial coefficients! The n-th row represents all the number of the form $\binom{n}{k}$, where the coefficient k is increasing when we read from left to right. There are plenty of interesting properties of Pascal's Triangle, and we will explore several of these in our examples and exercises.



Proposition 8.4.3 (Chairperson Identity)**.** *For any $n, k \in \mathbb{N}$,*

$$k\binom{n}{k} = n\binom{n-1}{k-1}$$

Proof strategy: Both sides of this equation are *products* of two terms, so we are looking for two two-step processes that construct the same set of elements. The term on the left can be thought of as $\binom{n}{k} \cdot k$, since multiplication is commutative. This indicates a selection of k items from $[n]$, and then ... something else. If we rewrite k as $\binom{k}{1}$, this makes the second step clear: we are then selecting one element from the k elements chosen in the first step.

This brings up a new strategy for describing subsets and particular elements contained therein: *committees* and *leaders*. This is a rather popular strategy in combinatorial proofs, because it reduces the amount of technical, mathematical language and symbols used, and makes for a better, colloquial understanding of the key ideas of the proof. We will show you how to use this strategy in the proof below, and then we will compare it to a proof that is written in more mathematical language. Before reading on, see if you can anticipate how we describe the right-hand side of the identity in question ...

Proof. Let $k, n \in \mathbb{N}$ be given, and let S be the set committees of k people, chosen from n people, with one designated committe chairperson. One way to construct elements of S is to first select a committee of k people, and then select one member of the committee to serve as the chairperson. By the Rule of Product, we conclude that

$$|S| = \binom{n}{k} \cdot \binom{k}{1} = k\binom{n}{k}$$

Another way to construct elements of S is to first select the committe chairperson from the entire pool of n people, and then select $k-1$ people from the remaining $n-1$ to fill out the commmittee. By the Rule of Product, we conclude that

$$|S| = \binom{n}{1} \cdot \binom{n-1}{k-1} = n\binom{n-1}{k-1}$$

By equating these two expressions for $|S|$, we conclude that

$$k\binom{n}{k} = n\binom{n-1}{k-1}$$

□

Proof summary: Count the k-committees from n people with a chairperson by selecting the committee and then the chair, or selecting the chair and the rest of the committee.

Comments: What if we had tried to describe this proof in purely mathematical terms, i.e. sets? What exactly does a "chairperson" correspond to in set-theoretical language? A committee of size k from n total people is just a set $T \subseteq [n]$ with $|T| = k$, but how do we distinguish this set from all of the k different ways to choose one of its members as the chairperson? One reasonable way is to define an *ordered pair* where the first coordinate is the set of committee members, and the second coordinate is the particular chairperson. With this strategy in mind for the proof, we would define

$$\hat{S} = \{(T, x) \mid T \subseteq [n] \wedge |T| = k \wedge x \in T\}$$

This set $\hat{S}$ is *equivalent* to the set S we defined in the proof above, in that it includes all of the ways to have a k-committee with one chairperson. In describing the two ways to count the elements of $\hat{S}$, though, we would likely find ourselves resorting to the same colloquial descriptions of committees and chairpeople! (Go ahead, try to count the elements of $\hat{S}$ *without* doing so.) It's just more natural and easier to understand that way. In short, there is no real reason to rigorously write out these set-theoretic descriptions of the sets of committees; however, it is important to point out that we *can*. This verifies that our descriptions in the above proof really are rigorous enough; they are rooted in mathematical concepts, but are easier to understand and follow when described in other terms.

Several examples of counting in two ways proofs involving committees and subcommittees are explored in the exercises of this section. We will present one more example here, as practice.

Proposition 8.4.4 (Committees of all Sizes)**.** *Let $n \in \mathbb{N}$. Then,*

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Proof strategy: The right-hand side could represent a number of things, but it seems likely that it involves an n-step process, where each step has 2 options. Let's re-examine that term in a minute. The left-hand side represents a *partition*, because we have a *sum* of several terms. Each individual term in the summation, of the form $\binom{n}{k}$, represents the number of ways to select a committee of k people, chosen from n total people. When we allow k to range from 0 to n, we are considering *all possible sizes* of committees. This indicates that we are counting *all* possible committees chosen from a set of n people. Now that we know what the right-hand side must count, we can construct an argument for that ... Try to do that on your own before reading our proof!

Proof. Let $n \in \mathbb{N}$ and let S be the set of all committees, of any size, chosen from a set of n people. Each element of S is a committee of a certain size from 0 to n, inclusive; for every $k \in [n] \cup \{0\}$, let S_k be the set of committees that have size exactly k. Then the set $\{S_k \mid k \in [n] \cup \{0\}\}$ is a partition of S. Thus, by the Rule of Sum, we conclude that

$$|S| = \sum_{k=0}^{n} |S_k| = \sum_{k=0}^{n} \binom{n}{k}$$

where $|S_k| = \binom{n}{k}$ because S_k is the set of all k-selections from $[n]$.

We can also count the elements of S as follows: take our set of n people and assign them numbers from 1 to n. (This can be done by giving everyone a t-shirt with their unique number on it, say.) To construct a committe, we line everyone up in numerical order and move along the line, saying "Yes" or "No" to each person, indicating whether or not they belong to the committee we are creating. Every sequence of n "Yes" and "No" assignments produces a unique committee. Since this is a n-step process with two choices at each step, the Rule of Product tells us there are 2^n ways to complete this process, so $|S| =^n$. By equating these two expressions for $|S|$, we conclude that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

□

Proof summary: Count all subsets of $[n]$ by partitioning based on size. (Note: this summary is written in terms of *sets*, but the proof itself is easier to write and understand in terms of *committees*, we feel.)

Maybe this proof seems a little verbose to you, particularly since we already proved that $|\mathcal{P}([n])| = 2^n$, by induction. Since we are considering *all* committees of all sizes, we are equivalently saying "Let $S = \mathcal{P}([n])$" and then counting S in two ways. However, when we start writing the proof in terms of *committees* we can't switch to talking about subsets of $[n]$ without writing a sentence or two about *why* those formulations are equivalent. As an exercise, try rewriting this proof entirely in set notation, without referencing committees. Which do you prefer?

The Summation Identity

The next combinatorial identity is very useful, and will appear in subsequent proofs and exercises in this chapter, so we present the result here. Furthermore, we will present *two different* counting in two ways proofs, and a third is covered in the exercises, even! These two proofs we present cover more standard counting objects, as they appear in counting in two ways proofs. We encourage you to read through both proofs and try to understand how they are *related*. Perhaps you're wondering why we would even bother to present two proofs of the *same* fact. ("Isn't one enough, so we already know it's true?") By understanding the proof structures and how they are *equivalent*, you will gain a deeper understanding of these proof techniques and be able to apply them better. Trust us! In addition, we will compare these techniques to the committees approach we used in the previous problem, and investigate out how all three methods are related.

Theorem 8.4.5 (Summation Identity)**.** *Let $n, k \in \mathbb{N}$. Then,*

$$\sum_{i=0}^{n} \binom{i}{k} = \binom{n+1}{k+1}$$

Proof 1 strategy: Seeing a single binomial term on the right-hand side indicates that we are looking at subsets of $[n+1]$ with size exactly $k+1$. With a summation on the left-hand side, we are partitioning the set of all such subsets based on some property. Since the binomial coefficient inside the summation has a k, instead of a $k+1$, in the bottom term, this means the index i somehow represents a particuar element being included in a subset. See if you can fill in the details of this partition before reading on ...

Proof 1. Let $n, k \in \mathbb{N}$ and define

$$S = \{T \subseteq [n+1] \mid |T| = k+1\}$$

By the definition of $(k+1)$-selections from $[n+1]$, we know that $|S| = \binom{n+1}{k+1}$. Next, for every $i \in [n] \cup \{0\}$, define the set

$$S_i = \{T \in S \mid i+1 \in T \wedge (\forall j \in T.\, j \leq i+1)\}$$

That is, S_i is the set of all subsets of $[n+1]$ with size $k+1$ whose *maximally-indexed* element is $i+1$. We claim that $\{S_i \mid i \in [n] \cup \{0\}\}$ is a partition of S.

First, observe that $S_i \cap S_j = \varnothing$ whenever $i \neq j$. This is because $T \in S_i$ implies $i+1 \in T$; further, if $i > j$ then the maximally-indexed element of any $U \in S_j$ is $j+1$ which is less than $i+1$, and if $i < j$ then any $U \in S_j$ contains $j+1$ but $j+1 \notin T$.

Second, notice that every $T \in S$ has some maximally-indexed element between 1 and $n+1$, and thus belongs to one of the S_i sets. (As a guide to this section of the proof, we have included a figure below that illustrates the case for $n = 4$

and $k = 2$. Notice that several of the sets are empty. In general, $S_i = \varnothing$ for every $i \in [k-1] \cup \{0\}$, but this makes sense because $\binom{i}{k} = 0$ for all of those values of i, as well.)

Next, we must identify $|S_i|$ for every $i \in [n] \cup \{0\}$. To construct an element $T \in S_i$, we identify a two-step process: (1) we include the element $i+1 \in T$, then (2) from the i smaller-indexed elements, we select k more to include. By the Rule of Product and the definition of selection, there are $\binom{i}{k}$ ways to do this.

Thus, by the Rule of Sum, we conclude that

$$|S| = \sum_{i=0}^{n} |S_i| = \sum_{i=0}^{n} \binom{i}{k}$$

By equating these two expressions for $|S|$, we conclude that

$$\sum_{i=0}^{n} \binom{i}{k} = \binom{n+1}{k+1}$$

□

Diagram for $n = 4$ and $k = 2$:

$$S = \Big\{ \{1,2,3\},\ \{1,2,4\},\ \{1,2,5\},\ \{1,3,4\},\ \{1,3,5\},\ \{1,4,5\},\ \{2,3,4\},\ \{2,3,5\},\ \{2,4,5\},\ \{3,4,5\} \Big\}$$

$$S_1 = \varnothing$$

$$S_2 = \varnothing$$

$$S_3 = \Big\{ \{1,2,3\} \Big\}$$

$$S_4 = \Big\{ \{1,2,4\},\ \{1,3,4\},\ \{2,3,4\} \Big\}$$

$$S_5 = \Big\{ \{1,2,5\},\ \{1,3,5\},\ \{1,4,5\},\ \{2,3,5\},\ \{2,4,5\},\ \{3,4,5\} \Big\}$$

Proof 1 summary: Count the $(k+1)$-element subsets of $[n+1]$ by partitioning based on the maximally-indexed element of any subset.

This proof strategy was developed from our initial observation that a binomial coefficient like $\binom{n+1}{k+1}$ represents a *selection*, identifying a subset of $[n+1]$. However, we could have interpreted this coefficient via another standard set of counting objects: binary tuples. This would lead to a different consideration of the summation on the left-hand side. Let's dive into that proof now!

Proof 2. Let $n, k \in \mathbb{N}$ and let S be the set of all binary $(n+1)$-tuples with exactly $k+1$ 1s. That is, $S \subset \{0,1\}^{n+1}$, and every $T \in S$ consists of exactly $k+1$ 1s and $(n+1)-(k+1) = n-k$ 0s.

We can identify $|S|$ directly by noting that constructing an element of S amounts to selecting, from $n+1$ open positions, $k+1$ positions to be filled with 1s (and then deterministically filling the remaining positions with 0s). Thus, $|S| = \binom{n+1}{k+1}$.

Next, we can identify a partition of S by classifying the tuples based on where the *right-most* 1 appears. Specifically, for $i \in [n+1]$, let S_i be the subset of S consisting of the tuples whose rightmost 1 occurs at position i, (as read from left to right, naturally). (See the diagram below the proof for an illustrated case with specific values of n and k.) To count the elements of S_i, we place a 1 in position i, then from the $i-1$ positions on the left, select k of them to be 1s. The remaining positions are deterministically filled with 0s. By the Rule of Product, $|S_i| = \binom{i-1}{k}$.

Now, we verify that $\{S_i \mid i \in [n+1]\}$ is a partition of S. First, observe that $S_i \cap S_j = \varnothing$ whenever $i \neq j$; if $i < j$ then the j-th position of any element of S_i is 0 but the j-th position of any element of S_j is 1, so any $(n+1)$-tuple cannot belong to both sets. Similarly, if $j < i$, the i-th position is either 1 (for elements of S_i) or 0 (for elements of S_j). Second, observe that any element of S has a rightmost 1, and that must occur in some position between 1 and $n+1$, so every element of S belongs to one of the S_i sets.

Thus, by the Rule of Sum

$$|S| = \sum_{j=1}^{n+1} |S_j| = \sum_{j=1}^{n+1} \binom{j-1}{k}$$

By redefining the index of summation as $i = j-1$, we can write

$$|S| = \sum_{i=0}^{n} \binom{i}{k}$$

and equating the two expressions for $|S|$, we have proved the result. □

Diagram for $n = 4$ and $k = 2$:

$$S = \Big\{ \{11100\},\ \{11010\},\ \{11001\},\ \{10110\},\ \{10101\},\ \{10011\},\ \{01110\},\ \{01101\},\ \{01011\},\ \{00111\} \Big\}$$

$$S_1 = \varnothing$$

$$S_2 = \varnothing$$

$$S_3 = \Big\{ \{11100\} \Big\}$$

$$S_4 = \Big\{ \{11010\},\ \{10110\},\ \{01110\} \Big\}$$

$$S_5 = \Big\{ \{11001\},\ \{10101\},\ \{10011\},\ \{01101\},\ \{01011\},\ \{00111\} \Big\}$$

Proof 2 summary: Count the binary $(n+1)$-tuples with exactly $k+1$ 1s by partitioning based on where the rightmost 1 occurs.

We hope that this has given you a good idea of how counting in two ways arguments are presented, and also how one tries to come up with such an argument by looking at the form of the identity. This subsection should give you some practice and let you attempt the exercises at the end of this section. If you find yourself needing more assistance, we suggest reading the next sections. It describes some heuristic methods to look at a counting in two ways problem and come up with an "appropriate" set S for a proof. These methods are based on the standard counting objects we presented earlier in this chapter, and their corresponding formulas.

Before we move on, though, we want to present one final counting in two ways proof because we find it extremely enlightening and clever and elegant. We don't expect you to come up with this kind of argument–particularly because it doesn't fit *exactly* into our description of "counting in two ways" proofs that we have developed thus far–but we think it is worth reading and marveling at, so please do so.

Proposition 8.4.6 (Gauss' Sum by Pairs)**.** *For any $n \in \mathbb{N}$,*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

Proof. First, observe that $\frac{n(n+1)}{2} = \binom{n+1}{2}$.

Now, consider a regular triangular array of dots consisting of $n+1$ rows, with k dots in the k-th row. The sum on the left represents the "area" of the first n rows of the array, i.e. the total number of dots in those n rows.

Next, we establish a bijection between those dots, and *pairs* of dots in the $(n+1)$-th row. For any pair of dots, draw inwardly-pointing diagonal lines upwards

through the array to obtain a unique dot in an above row. Conversely, for any dot in the array, draw outwardly-pointing diagonal lines downward through the array to obtain a unique pair of dots in the bottom row. Thus, the number of pairs of dots in the $(n+1)$-th row, which is $\binom{n+1}{2}$, is equal to the number of dots in the upper array, which is $\sum_{k=1}^{n} k$. □

8.4.3 Standard Counting Objects

We have already discussed several standard combinatorial objects in the previous section. One of the difficulties of Counting in Two Ways proofs, though, is figuring out which objects to count! These exercises are very often posed as follows: "Here's an identity; prove it by Counting in Two Ways." This doesn't give you any idea of what to count, just that you need to count something! In this short section, we will do our best to provide a handy guide to "unraveling" a combinatorial identity and creating a Counting in Two Ways proof. These ideas are based on our experiences and some standard arguments used by combinatorialists.

Binomial Coefficients and Multiple Interpretations

These objects and any corresponding counting formulas were covered in the previous section, so we encourage you to reread any part of that section that feels unfamiliar. What we can do here is emphasize when to *recognize* that a certain counting object is somehow "relevant" to a counting in tow ways problem. For instance, recall the Chairperson Identity, but pretend we haven't proved it yet:

$$k\binom{n}{k} = n\binom{n-1}{k-1}$$

Seeing that the identity only contains products of binomial coefficients (and remembering that we can always write k as $\binom{k}{1}$, for instance), indicates that we should try to count something that is easily describable by simple binomial coefficients. The most natural choice is subsets of $[n]$; equivalently, we could use committees of a certain size chosen from a set of people, or binary n-tuples with k 1s. Any of these three choices would provide us with fairly easy descriptions of the individual terms in the expression and allow us to relate them. At that point, we need to choose which interpretation we feel most comfortable with, the one with which we can most easily explain all of the terms.

If we choose to use people and committees, then we can follow the argument we used in the proof above. If we opt for the subsets of $[n]$, then we need to come up with a reasonable two-step process to describe the product of terms on both sides of the identity. After choosing a subset of size k, what might the $\binom{k}{1}$ term represent? In essence, we are singling out a "special" element of the already chosen subset. This is similar to the $\binom{n}{1}$ term on the right-hand side, where we single out a "special" element first, and then fill in the rest of the subset. However, when our context is subsets of $[n]$, we no longer have the terminology "chairperson of the committee'. (This is why we feel

the committees interpretation is reasonable and easy to use. Just number all of the people uniquely, and then we can safely use any of this terminology.) Common interpretations of this term might involve, say, "circling" an element, to indicate it is special. That is, both sides of the equation would count subsets of $[n]$ with size k with one element of the subset circled. On the left-hand side, we choose the subset and then assign the circle; on the right-hand side, we circle an element and include it in the subset, then fill out the rest of the subset. There are other simple and understandable ways to accomplish this argument, but we wanted to make sure to point out that the terminilogy of "committees" doesn't apply unless we choose that setting from the beginning of the proof. (Challenge problem: how would you approach this proof in the context of binary n-tuples? Hint: think about alowing that "special" position to be filled by some symbol other than a 0 or a 1.)

It is also common to find a product of binomial coefficients where the "top terms" are identical. For instance, think about how to describe a term like the following, in the context of a counting in two ways proof. (Pretend it is just one side of an equation; the other is irrelevant for this discussion.)

$$\binom{n}{k}\binom{n}{\ell}$$

There are two reasonable ways to describe this type of product, and deciding which one to use will depend on the other side of the equation, or the other terms involved. We will present both interpretations here and let you figure out which one to use by examining the context.

Consider the committees context, so each term somehow represents choosing a committee of a certain size (k or ℓ) from n total people. One interpretation is that we choose two committees from the *same* set of n total people. That is, perhaps we have n professors in a department and we need to choose k of them to oversee the budget and ℓ of them to oversee the curricula, and professors are allowed to possibly serve on both committees. Another interpretation is that we choose two committees from *different* sets of people, but both of those sets have size n. That is, perhaps we have n boys and n girls in a class, and we want to choose k of the boys and ℓ of the girls to form a club. Either of these interpretations is "correct" and reasonable to use, but the "right choice" will certainly depend on the rest of the problem in question.

One final term that is useful in committee-type arguments is the idea of a subcommittee. Since a committee of k people, chosen from n people, already represents a *subset*, a subcommittee really represents a subset of a subset. Thus, if we find a term like

$$\binom{a}{b}\binom{b}{c}$$

in an identity, we might choose to interpret this as choosing a committee of b people from a total people, and *then* choosing a subcommittee of c people from those b people. This can be described as choosing a club and then its officers, or a sports team and then its starting squad, or anything like that.

Exponents and Processes

Other terms, beside binomial coefficients, that appear frequently in combinatorial identities are exponential terms: n^3, 2^n, n^{k-1}, and so on. Oftentimes, the interpretation for these terms will already be dictated by how one has assigned a context, based on the other terms in the identity. What we present here are some standard, common, and easily-explainable ways to interpret terms like this. What's interesting is that the interpretation sometimes might depend on whether the base or the exponent is the larger number!

Consider a term like

$$\binom{n}{k}2^k$$

Let's assume that we have assigned a "committees" interpretation to the problem, based on the rest of the identity, and that we have declared that the binomial coefficient $\binom{n}{k}$ represents a selection of a committee of k people from a class of n students. What does the 2^k term then represent? Remember that this term could come from a k-step process, where each step can be done in one of two ways. Since k is the size of the committee chosen, then we can simply describe some 2-step decision process for every member of the committee. For instance, we could assign every committee member either a red hat or a blue hat; or, for each committee member, we could choose whether or not to give him/her a gold star; or, we could force the committee members to choose whether to become a Republican or a Democrat. Be creative, if you'd like! Of course, the chosen interpretation will have to work with the rest of the identity, so sometimes one interpretation is easier to explain than the other. Keep that in mind, and be willing to go back and change your interpretation if you find that it's hard to convey what you're thinking.

Now, consider a term like

$$\binom{n}{k}2^n$$

Again, assume we have assigned a "committees" interpretation to the problem. How is this different from the above situation? In this case, the exponent matches the "top term" of the binomial coefficient. Thus, the selection of a committee doesn't necessarily have anything to do with the subsequent n-step process. This term might describe a selection of k class officers from an n student class, and then an assignment of every student into either Block A or Block B (regardless of the officer assignments). If we *weren't* using the "committees" interpretation, this term might describe a binary n-tuple with exactly k 1s where some unknown number of the 0s and 1s are circled. The choice of interpretation will depend on the rest of the problem and how comfortable you feel explaining the terms.

Think about how to modify these interpretations with slightly different numbers. With a term like

$$\binom{n}{k}4^n$$

we might describe a committee of k members, each of whom has a Red, Blue, Green, or Yellow hat. With a term like

$$\binom{n}{k}5^n$$

we might describe a binary n-tuple with exactly k 1s, where every 0 and 1 has either 1, 2, 3, 4 or 5 circles around it.

Next, let's examine some terms where the base is a variable and the exponent is fixed. For instance, consider a term like

$$\binom{n}{k}k^2$$

In this case, we are somehow selecting k objects from n total objects and performing a 2-step process with k choices at each step. That is, the selection of the k objects first affects the outcome of the second part, the 2-step process. If we are operating in a "committees" context, we might interpret this term as selecting a committee of k people, and then choosing 2 officers of the committee–say, a speaker and a treasurer–where any person on the committee can be chosen as an officer and, furthermore, any member can potentially serve in both office positions.

If we wanted to use a "tuples" interpretation, we might describe this term as selecting a binary n-tuple with exactly k 1s, where one 1 has a circle around it and one 1 has a box around it (and it's possible that the same 1 has a circle and a box). We also might describe this term using an "alphabet" interpretation. From an alphabet of n total letters, we can choose a subset of k letters and then construct two-letter words from just those k letters. Think about each of these three interpretations and why they all work, and how they relate to each other. Try to take one of our proofs and rewrite it using each of these interpretations. Also, think about how these interpretations would be different with a k^3 or k^4, say.

Now, consider a term like

$$\binom{n}{k}n^3$$

in a "committees" context. Since the base of the exponential term is the same as the "top term" of the binomial coefficient, there isn't necessarily a relationship between the committee that the $\binom{n}{k}$ term represents and the subsequent 3-step process. Thus, we might describe a selection of a k-person committee, and then an assignment of one Red, one Blue, and one Green ribbon, where one person might receive multiple ribbons, and anyone (on or off the committee) may receive a ribbon (or ribbons). We will leave it to you to construct an appropriate interpretation of a term like this under a "binary n-tuples" interpretation. Do try it!

Summation means Partition

It is quite common to find a *summation* in a combinatorial identity. Handling this in a counting in two ways proof is a little more intricate because a summation represents several terms at once. The most important rule, though, is this: a summation represents a *partition*. Always. In particular, it represents a partition and tells us what the cardinalities of all of the partition sets are. To explain this in a counting in two ways proof, there are always three properties we need to describe:

- What the sets of the partition are.
- Why the *limits* on the index of the sum make sense in the context.
- For an arbitrary index, why the size of the corresponding set is the term in the sum.

We will illustrate those through an example.

Example 8.4.7. **Pro/Con Committee Identity:**

$$\binom{n}{k}2^{n-k} = \sum_{i=k}^{n}\binom{n}{i}\binom{i}{k}$$

Intuition: Create a committee of size k from n people. Then, determine whether the non-committee people are for or against the committee's decisions. We could also do this by first selecting at least k people who will be on/for the committee, and set everyone else to be off and against. Then, from that pool, we select k people to actually be on the committee, setting the others to be for it. (Note: It's important to say all of the steps we will perform in these constructions. Don't assume anything is obvious to the reader.)

Proof. Consider a set of n people. Let S be the set of ways to select k of the n people to be on a committee, with every person who isn't on the committee having a firm opinion to be For or Against the committee.

First, we can find $|S|$ by a multi-step process:

- Select k of the n people to be on the committee.

  $\mathbf{\binom{n}{k}}$ **options**

- For each of the remaining $n-k$ people, have them decide whether to be For or Against. This is a process with $n-k$ steps and two choices at each step, so by ROP ...

  $\mathbf{2^{n-k}}$ **options**

By ROP, we have $|S| = \binom{n}{k}\cdot 2^{n-k}$.

Second, we can find $|S|$ by establishing a partition based on how many people are

For the committee. By the definition of S, anywhere from none to all of the $n-k$ non-committee members could be For the committee. In total, then, between the committee members and their For supporters, we can have somewhere from k to $k + (n - k) = n$ people, inclusive.

For each i that satisfies $k \leq i \leq n$, let $S_i \subseteq S$ be the set of ways to have k committee members and $i - k$ For supporters. (Notice that $0 \leq i - k \leq n - k$, which matches the restriction noted previously.)

Notice that $\{S_i \mid k \leq i \leq n\}$ is a partition of S. This is because, for any element of S, that element can be characterized by how many For supporters of the committee there are, and that has to be some specific number. Now, we can find $|S_i|$, for each such i, by a multi-step process:

- From all n people, select i people. These are potential committee candidates.

  $\binom{\mathbf{n}}{\mathbf{i}}$ **options**

- Designate the $n - i$ *other* people to be decidedly Against the committee we are constructing.

  (This step is deterministic, so there is only 1 way to do it, but we need to point this out to fully describe an outcome that is an element of S.)

- Of those i people chosen in the first step, select k to be actual committee members.

  $\binom{\mathbf{i}}{\mathbf{k}}$ **options**

- Designate the $i - k$ people not chosen in the previous step to be For supporters of the committee, but not committee members.

  (Again, there is one way to do this, but it is required to fully describe the outcome.)

By ROP, we find that $|S_i| = \binom{n}{i}\binom{i}{k}$.

By ROS, then, we find that $|S| = \sum_{i=k}^{n} |S_i| = \sum_{i=k}^{n} \binom{n}{i}\binom{i}{k}$.

Since we found $|S|$ in two ways, we can equate them. This proves the claim. □

Notice that we did several things after identifying the partition in our proof. We explained why it is a partition. We explained how it was related to the index on the summation. We explained how the *limits* on the sum correspond to the partition and represent all possibilities. Then, for an arbitrary i, we explained why $|S_i|$ is the corresponding term in the sum.

8.4.4 Binomial Theorem

We can prove a powerful and important theorem using this proof technique, counting in two ways. This will be an interesting application of this technique, but this is also a useful result in its own right, as we will see!

Theorem 8.4.8. *Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$. Then,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

We will explain a few different ways to prove this.

Proof 1. Consider proving this where we assume $x, y \in \mathbb{N}$.

In this case, think of having a set of $x + y$ symbols; for instance, say we have x lowercase letters and y capital letters. Then $(x + y)^n$ is the number of strings of length n made from those symbols.

On the right-hand side, we partition the set of all such length n strings based on how many positions in a string are filled with lowercase letters. There will be somewhere from 0 to n of the positions (inclusive) filled with choices from the set of x lowercase letters. For each such k, with $0 \leq k \leq n$, the number of length n strings with exactly k lowercase letters is $\binom{n}{k} \cdot x^k \cdot y^{n-k}$ since we select those k positions for the lowercase letters, choose how to fill those positions, and then fill the remaining $n - k$ positions with capital letters. □

Proof 2. Let's prove this for general $x, y \in \mathbb{R}$ by counting the number of terms in the "FOILed" expansion that correspond to k choices of x (and thus $n - k$ choices of y) from the factors in the product.

Consider the product

$$(x+y)^n = \underbrace{(x+y) \cdot (x+y) \cdots (x+y)}_{n \text{ factors}}$$

Think about multiplying out these n factors by applying the Distributive Property over and over. For example, with $n = 2$, we have

$$\begin{aligned}(x+y)^2 &= (x+y)(x+y) = x(x+y) + y(x+y) = x \cdot x + x \cdot y + x \cdot y + y \cdot y \\ &= x^2 + 2xy + y^2\end{aligned}$$

and with $n = 3$, we have

$$(x+y)^3 = (x+y)(x+y)(x+y) = x(x+y)(x+y) + y(x+y)(x+y) = \cdots$$

The general idea is this: To find a term in the ultimate product, we select either an x or a y from each factor $(x + y)$. Every such term looks like $x^k \cdot y^{n-k}$, for some k between 0 and n. All we need to do is identify how many *ways* there are to create a term like $x^k \cdot y^{n-k}$. This amounts to finding the number of ways to select k of the n factors, and say that we chose "x" from those factors and "y" from the other $n - k$ factors. By definition of selection, there are precisely $\binom{n}{k}$ many ways to do this! □

Proof 3. We could also prove this by induction! **Pascal's Identity** is essential in the Induction Step. This is presented to you in Exercise 8.9.14. □

Example 8.4.9. Let's show how this theorem is useful.

- Apply the Binomial Theorem to show

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

Proof. Use $x = 1$ and $y = 1$. □

That's it! A result we proved already by induction and then by a counting in two ways argument now follows *immediately* from this powerful theorem.

- Prove that the number of odd-sized subsets of $[n]$ is equal to the number of even-sized subsets of $[n]$; that is,

$$\sum_{k=0}^{\lceil n/2 \rceil - 1} \binom{n}{2k+1} = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k}$$

One can prove this by finding a bijection between the set of even-sized subsets and the set of odd-sized subsets. We could even try to explain this with a counting argument.

Instead, let's subtract on both sides and rewrite the equality as

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$$

Notice that this is precisely what Binomial Theorem says, with $x = -1$ and $y = 1$. Amazing!

8.4.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the overall method of a **counting in two ways** argument?

(2) Write a short *proof summary* for each example proof of this section.

(3) When there is a summation in a propose identity, what must we discuss in a subsequent counting in two ways proof?

(4) What are the different ways that we proved the Summation Identity? How are they fundamentally the same?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Let $\ell, k, n \in \mathbb{N}$ be given. Prove that

$$\binom{n}{k}\binom{k}{\ell} = \binom{n}{\ell}\binom{n-\ell}{-\ell}$$

by a counting in two ways argument.

(2) Prove that

$$n \cdot 2^{n-1} = \sum_{k=1}^{n} \binom{n}{k} \cdot k$$

by a counting in two ways argument.

(3) Prove

$$3^n = \sum_{k=0}^{n} \binom{n}{k} 2^{n-k} = \sum_{k=0}^{n} \binom{n}{k} 2^k$$

by a counting in two ways argument.

(**Hint:** Consider using the set of ternary strings.)

Then, explain how it follows from the Binomial Theorem, as well.

(4) Prove that $k^2 = \binom{k}{1} + 2\binom{k}{2}$ by a counting in two ways argument.

Apply the **Summation Identity** to deduce that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

(5) Prove the following **Geometric Series Formula** by a counting in two ways proof:

$$\forall q \in \mathbb{N} - \{1\}\,.\ \forall n \in \mathbb{N}\,.\quad 1 + q + q^2 + q^3 + \cdots + q^{n-1} = \sum_{k=0}^{n-1} q^k = \frac{q^n - 1}{q - 1}$$

(Note: This formula holds true, in fact, for any *real* number $q \neq 1$, but the counting in two ways proof we are asking for only applies to *natural* numbers $q \neq 1$. To prove the real-valued version, use induction.)

(**Hint:** Consider the set of all n-tuples made from q elements, except for a particular one...)

8.5 Selections with Repetition

8.5.1 Motivation

When we derived formulas for the number of *arrangements* and *selections*, we took care to point out whether we are allowed to **repeat** objects. At the time, we left off deriving a formula for the number of ways to select objects *with repetition*. Specifically, we needed to develop—and become comfortable with—the technique of *counting in two ways* before tackling this problem. Now, we're ready!

Example 8.5.1. Say I have a box of fruit on my kitchen counter. It has a bunch of apples and bananas and peaches in it. Let's say at least 10 of each. I reach in and grab 5 pieces of fruit to have throughout my day at school. How many different combinations could I have possibly grabbed?

For the sake of this example, we are assuming that any two apples, for instance, are *indistinguishable* to me. None of them are off-colored or far smaller than the others, or anything like that. With this assumption, this question is about **selecting** 5 objects from 3 **types** of objects. The outcome is *unordered* (so it's a selection, not an arrangement) and we are allowed to *repeat* objects from any type (i.e. I can pick several bananas).

For example, we could pick 4 apples and 1 peach, or 5 bananas, or 1 apple and 2 bananas and 2 peaches.

This represents the most general form of this kind of problem. Given a number of **types**, how many ways can I select some number of total objects from those types? Let's see another example before finding a formula for these kinds of problems. In fact, this is the interpretation that we will use to derive the formula!

Example 8.5.2. Suppose we have n indistinguishable (identical) gold coins to distribute amongst k distinguishable (distinct and labeled) pirates. How many ways can we do this?

Try working with this for small values of n and k. Actually, grab some nickels and some friends, and try to work it out. If you have $n = 5$ coins and $k = 3$ friends, how many ways can you distribute the coins?

Keep in mind that the pirates are *distinguishable*. For instance, giving Captain Redbeard 2 coins and Captain Blackbeard 10 coins is *not* the same outcome as giving Redbears 10 and Blackbeard 2. We should count those separately.

Also, keep in mind that the coins are *indistinguishable*. This means that it doesn't matter *how* we hand out the coins, or in what *order*, or anything like that. All that matters is the final outcome, where they all end up. For instance, giving Redbeard 5 coins, then Blackbeard 5, then Redbeard another 2 . . . that's the same as just giving Redbeard 7 coins and Blackbeard 5. We should count these as the same outcome.

Try to work with these examples and come up with a formula that solves these problems. Can you generalize to any n and k? Can you prove your claim?

Try it! Then, read the next section for our formula and proof.

8.5.2 Formula

We will derive a formula for the number of selections with repetition by considering the Pirates & Gold example. First, let's explain why this is like selection with repetition:

Pretend that we are the *gold distributor* in this scenario. We are sitting at a table with k pirates seated around us and a bag full of n gold coins at our side. We can choose how to distribute the coins by passing them out one by one. When we choose to give a gold coin to Pirate $\#i$ (for some $i \in [n]$), we are *selecting* that pirate from amongst the k different *types* of pirate. Ultimately, to pass out n pieces of gold, we need to make n *selections*. Thus, we are selecting n objects from k types, overall, and we are allowed to *repeatedly* select a type.

Derivation

Think of having our n coins laid out on the table, all in a row. To distribute them amongst the k pirates, we need to lay down "dividers" or "bars" that separate the gold pieces into k piles. Then, Pirate 1 will take the pile on the left, Pirate 2 will take the next pile, etc.

This "dividers" argument allows us to count the number of ways to complete this process conveniently! To assign the n pieces of gold to k distinct pirates, we need to have n coins separated by $k-1$ dividers. Think about why we only need $k-1$ dividers here. It's easy to see why we only need 1 divider to split a row of coins into 2 piles. Then, we can see that 2 dividers split them into 3 piles. In general, once we have already laid down $k-1$ dividers, there are k piles established; we don't need to lay down a final divider at the *end* of the row to represent that the rightmost pile goes to Pirate $\#k$.

Example 8.5.3. For example, with $k = 3$ and $n = 7$, we might have a distribution like this:

$$\circ \mid \circ\circ\circ \mid \circ\circ$$

In this case, Pirate #1 receives 1 gold, Pirate #2 receives 4 gold, and Pirate #3 receives 2 gold.

Notice that this is *different* from the following outcome:

$$\circ\circ\circ \mid \circ\circ \mid \circ$$

In this case, Pirate #1 receives 4 gold, Pirate #2 receives 2 gold, and Pirate #3 receives 1 gold.

We could also have some pirates receive 0 gold:

$$\circ\circ\circ\circ\circ \mid \circ\circ \mid$$

Here, Pirate #1 receives 5 gold, Pirate #2 receives 2 gold, and Pirate #3 receives 0 gold.

What do these observations tell us? Well, this means that any assignment of the gold pieces corresponds to a string of length $n+k-1$ with exactly n coins and exactly $k-1$ dividers. There is a *bijection* between the sets of these two objects: gold distributions and divider placements. Given a gold distribution scheme, we can construct the corresponding divider arrangement. (For instance, if we were told Pirate 1 is to receive 5 gold, Pirate 2 is to receive 2 gold, and Pirate 3 is to receive 0 gold, we would construct the divider arrangement in the last example above.) Likewise, given a divider arrangement, we can read it off and determine what gold distribution scheme it corresponds to.

This bijection tells us that to count the number of ways to distribute the gold, we just need to count the number of possible divider arrangements there are. We can count these quite easily! A divider arrangements is just a string of length $n+k-1$ with exactly $k-1$ dividers. This is because we need n gold and $k-1$ dividers, so $n+k-1$ positions, in total. Thus, by the definition of selection, there are

$$\binom{n+k-1}{k-1}$$

ways to construct such an arragement!

(You might have also heard of this argument as "Stars and Bars". This is just another common interpretation of this problem, where the gold pieces are replaced with Stars and the dividers are replaced with Bars.)

Because of that bijection between the set of such arrangements and the set of gold assignments, we conclude that $\binom{n+k-1}{n}$ is the number of ways we can distribute the gold!

We already know that $\binom{n}{k} = binn-k$, in general, so we could apply that here and deduce that the number of gold assignments is also

$$\binom{n+k-1}{n}$$

But we could already have seen this in our derivation. We need to construct a string of length $n+k-1$ with $k-1$ dividers (and the remaining positions being gold pieces). Equivalently, we need a string of that length with n gold pieces (and the remaining positions being dividers).

8.5.3 Equivalent Forms

Before moving on to solve some problems with this new formula, let's consider some **equivalent formulations** of a fundamental *selections with repetition* problem. Whenever you face a problem that involves these concepts or formulations, you might consider applying the formual we just derived, somehow.

Pirates & Gold

This is the original formulation we used to derive the formula, so certainly it is applicable in a context like this. In general, all we need to know is the number of pirates and the number of gold pieces.

> The number of ways to distribute n identical pieces of gold amongst k distinguishable pirates is $\binom{n+k-1}{k-1}$.

Implicit in our derivation, mind you, is that pirates could conceivably receive 0 gold pieces, so keep that in mind. Some problems might ask you to consider other *conditions* on the distributions. For example, what if every pirates must receive *at least one* piece of gold?

Integer Sums

Consider reformulating the Pirates & Gold problem as follows. Let's define $x_i \in \mathbb{N} \cup \{0\}$ to be the number of gold pieces that Pirate $\#i$ receives in the distribution. The conditions of the problem require that

$$\forall i \in [k] \text{.}\ x_i \in \mathbb{N} \cup \{0\}$$

and that

$$\sum_{i=1}^{k} x_i = x_1 + x_2 + x_3 + \cdots + x_k = n$$

Aha! What if we had asked about the number of **solutions** to such an equation? This corresponds exactly (in a *bijective* manner) with the ways to solve the Pirates & Gold problem. Given a solution to this equation, we just give Pirate $\#i$ exactly x_i pieces of gold. This gives us a different way of stating the problem:

> The number of solutions to the equation $x_1 + x_2 + \cdots + x_k = n$ that satisfy the condition $\forall i \in [k] \text{.}\ x_i \in \mathbb{N} \cup \{0\}$ is $\binom{n+k-1}{k-1}$.

Balls and Bins

What if we were given n identical balls, and we were asked to place them in k different bins. (The bins are distinguishable, so let's say they are labeled from 1 to k.) How many ways can we do this? This is easy to relate to the previous formulation! Let $x_i \in \mathbb{N} \cup \{0\}$ be the number of balls that end up in Bin $\#i$. Then the same exact conditions as the problem above apply here.

The number of ways to distribute n identical balls into k distinguishable bins is $\binom{n+k-1}{k-1}$.

Indistinguishable Dice

Consider rolling n *identical* dice. How many outcomes are there? This is **not** the same as rolling distinguishable dice (where the dice are different colors, for example.) Instead, an outcome of this process is an *unordered* list of the numbers that are showing on the faces.

For example, if we rolled 3 indistinguishable 6-sided dice, an outcome of that process might be the **unordered** list $(1, 3, 3)$. To think about this, pretend your friend went into another room and rolled 3 dice, then came back and told you what happened. If he says "I rolled a 1 and two 3s", then you didn't learn about *which* dice showed which number. (Contrast this with the case where he says "I rolled a 1 on the first die and then a 3 on each of the second and third dice.") By asking about the number of outcomes of indistiguishable dice, we are essentially asking about how many possible responses your friend could give you that do not indicate anything about the *order* in which the rolls appeared.

We can relate this to the "Balls and Bins" formulation by rolling all the dice and placing them into 6 numbered boxes based on what the faces show. Equivalently, to characterize an outcome of this process, we need to know how many dice showed 1, how many showed 2, etc. We don't care (nor could we know!) *which* dice showed which numbers; we only need to know *how many* showed each number.

The number of outcomes of rolling n indistinguishable k-sided dice is $\binom{n+k-1}{k-1}$.

8.5.4 Examples

Let's practice using this newly-derived formula to solve some problems! We'll examine a couple of different formulations of the fundamental result, in the process.

Example 8.5.4. Let's say we have $n = 20$ pieces of gold to distribute amongst $k = 3$ pirates. Let's say the pirates are Captain Redbeard (Khair ad Din, an Ottoman), Captain Blackbeard (Edward Teach, an Englishman), and Captain Kidd (a Scotsman).

Let's figure out how many ways there are to distribute the gold under certain conditions:

(1) How many ways are there total?

This is like selecting 20 objects from 3 types, with repetition allowed. Whenever we select a pirate, that means we give him a piece of gold.

By the above selection formula, there are

$$\binom{20+3-1}{20} = \binom{22}{20} = \frac{22 \cdot 21}{2} = 231$$

ways to do this.

(2) How many ways ensure every pirate gets at least 2 pieces?

Let's just give everyone two pieces of gold right away. Then, we have $20-6 = 14$ pieces of gold left to distribute amongst all 3 pirates, so there are

$$\binom{14+2}{14} = \binom{16}{14} = \frac{16 \cdot 15}{2} = 120$$

ways to do this. Think about why this works. We are essentially re-defining what "getting 0 gold pieces" means. Instead of starting with 20 pieces and worrying about whether or not everyone gets *at least* two pieces, we can just ensure that condition is met right away, and then distribute what's left over.

(3) How many ways ensure Redbeard and Blackbeard get at least 2 and Kidd gets at least 6?

Just like the last one, let's give Redbear and Blackbeard 2 pieces each, and let's give Kidd 6 pieces. This leaves us with $20 - 4 - 6 = 10$ pieces left to distribute amongst all 3 pirates, so there are

$$\binom{10+2}{10} = \binom{12}{10} = \frac{12 \cdot 11}{2} = 66$$

ways to do this.

(4) How many ways ensure Redbeard and Blackbeard get at least 2 while Kidd gets no more than 2?

There are a couple of ways to approach this one.

(i) Let's establish cases based on whether Kidd gets 0 or 1 or 2 golds. In each case, we will give Redbeard and Blackbeard 2 pieceseach right away, and then give Kidd the corresponding amount (0 or 1 or 2). This leaves us with 16 or 15 or 14 pieces left to distribute *only* amongst the first two pirates, so there are

$$\binom{16+1}{16} + \binom{15+1}{15} + \binom{14+1}{14} = 17 + 16 + 15 = 48$$

(ii) Let's consider *all* of the ways to ensure Redbeard and Blackbeard get at least 2 golds each, and then *remove* from this the number of ways where Kidd gets too many, i.e. at least 3 golds.

If we give Red/Blackbeard 2 golds each, then we have 16 pieces left to distribute amongst all 3 of the pirates, so there are

$$\binom{16+2}{16} = \binom{18}{16} = \frac{18 \cdot 17}{2} = 153$$

ways to do this step.

Then, if we give Red/Blackbeard 2 golds each and give Kidd 3 pieces and then distribute the remaining 13 amongst all 3 of the pirates, there are

$$\binom{13+2}{13} = \binom{15}{13} = \frac{15 \cdot 14}{2} = 105$$

ways to do this step. We want to *remove* these possiblities from the previous count. Thus, in total, there are

$$\binom{18}{16} - \binom{15}{13} = 153 - 105 = 48$$

ways to give Redbeard and Blackbeard at least two each, but give Kidd no more than 2.

(Look, we get the same answer both ways!)

Example 8.5.5. Consider the following equation:

$$x_1 + x_2 + x_3 + x_4 + x_5 = 25$$

Let's identify the number of solutions to this equation where each variable x_i is a non-negative integer. We'll impose certain conditions and count the number of solutions that satisfy them.

(1) How many solutions are there total?

Applying the formula we derived, we see there are

$$\binom{25 + (5-1)}{5-1} = \binom{29}{4}$$

(2) How many solutions satisfy $x_1 \geq 4$?

This is exactly like asking for Captain Redbeard to get *at least* 4 gold pieces. We are going to "pre-distribute" 4 "counts" to the variable x_1, and then distribute the remaining 21 "counts" to all five variables.

More formally, we define $y_1 = x_1 - 4$. The condition requires only $y_1 \geq 0$. Thus, we are trying to solve the equation

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 = 25 &\iff (x_1 - 4) + x_2 + x_3 + x_4 + x_5 = 21 \\ &\iff y_1 + x_2 + x_3 + x_4 + x_5 = 21 \end{aligned}$$

Applying the formula, we see there are

$$\binom{21+(5-1)}{5-1} = \binom{25}{4}$$

such solutions.

(3) How many solutions satisfy $x_1, x_2 \geq 5$ and $x_3, x_4, x_5 \geq 2$?

Using a method exactly like the last one, we see that we're trying to solve

$$y_1 + y_2 + y_3 + y_4 + y_5 = 9$$

where $y_i = x_i - 5$ for $i = 1, 2$, and $y_i = x_i - 2$ for $i = 3, 4, 5$. The right-hand side has been changed to $25 - 5 - 5 - 2 - 2 - 2 = 9$. By the formula, we know there are

$$\binom{9+(5-1)}{5-1}$$

such solutions.

(4) How many solutions satisfy $x_2 \leq 5$?

We can do this in one of two ways. First, let's take the total number of solutions (found in the first part of this example) and *remove* the number of solutions that *fail* this desired condition. That is, let's take the number of solutions where $x_2 \geq 6$, which is

$$\binom{(25-6)+(5-1)}{5-1} = \binom{23}{4}$$

and remove it from the total, yielding

$$\binom{29}{4} - \binom{23}{4}$$

Second, we could write this as a sum, finding the number of solutions that satisfy $x_2 = \ell$, for $0 \leq \ell \leq 5$:

$$\sum_{\ell=0}^{5} \binom{28-\ell}{3} = \binom{28}{3} + \binom{27}{3} + \binom{26}{3} + \binom{25}{3} + \binom{24}{3} + \binom{23}{3}$$

Interestingly enough, if we had only thought to solve this in the second way, we could still *reduce* the expression to the first one. We just need to use the Summation Identity! Observe that

$$\begin{aligned}\sum_{\ell=0}^{5} \binom{28-\ell}{3} &= \binom{28}{3} + \binom{27}{3} + \binom{26}{3} + \binom{25}{3} + \binom{24}{3} + \binom{23}{3} \\ &= \sum_{k=0}^{28} \binom{k}{3} - \sum_{k=0}^{22} \binom{k}{3} \\ &= \binom{29}{4} - \binom{23}{4}\end{aligned}$$

Neat, right?

8.5.5 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What is the difference between a selection and a selection with repetition?

(2) What is the number of ways to select n objects from k objects? (Careful about letters here!)

(3) What is the number of ways to select n objects from k types of objects?

(4) How are the "Pirates & Gold" and "Integer Sums" formulations *equivalent*?

(5) Adapt the argument we used to derive the formula $\binom{n+k-1}{k-1}$ and use it to prove the same formula counts the number of solutions to the "Integer Sums" formulation.

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) A clothing store makes 5 different colors of shirts (red, green, blue, white, and black).

   (a) We need to buy 10 shirts. How many ways can we do this, assuming we can order any number of shirts of each color (i.e. there's an unlimited supply of each color)?

   (b) We need to acquire some shirts, either 10 or 11 or 12, we're not sure yet. How many ways can we do this, again assuming an unlimited supply of each color?

   (c) We need to acquire 10 or 11 or 12 shirts, but in each case we need at least 1 shirt of each color. How many ways can we do this?

   (d) Now, we need to order 25 shirts, but we are told that there are only 3 red shirts left (while each other color still has an unlimited supply). How many ways can we do this?

   (e) Now, we still need to order 25 shirts, but we are told there are only 3 red shirts and 5 blue shirts left (while each other color still has an unlimited supply). How many ways can we do this?

(2) Consider rolling 20 indistinguishable dice.

   (a) How many total outcomes are there?

   (b) How many have each number appearing at least twice?

   (c) How many have at most three 6s?

   (d) How many have at least four 6s?

(3) What is wrong with the "proof" of the following claim?

> Consider 4 buckets of coins: one bucket contains pennies, one contains nickels, one contains dimes, and one contains quarters. There are over 50 coins in each bucket (so we don't have to worry about running out of any type of coin).
>
> We want to select 50 coins from these buckets; we want to make sure that we select at least 10 pennies and at least 10 nickels but at most 10 dimes and at most 10 quarters.
>
> **Claim:** The number of ways to do this is
>
> $$\binom{53}{3} - \binom{11}{3} = 23261$$
>
> *Proof:* Consider the total number of ways to select 50 coins from 4 types, with no added restrictions. This is selecting $k = 50$ objects from $n = 4$ types, so there are $\binom{53}{3}$ ways to do this.
>
> Now, we want to remove from this total the number of ways to select the coins where we do pick at least 10 pennies and at least 10 nickels but also at least 11 dimes and at least 11 quarters. To count these selections, we just want to actually select all of those coins (there are 42 total) and then select 8 more from all four types. We know there are $\binom{11}{3}$ ways to choose $k = 8$ objects from $n = 4$ types.
>
> Subtracting the second number from the first, we get the number given in the claim.

8.6 Pigeonhole Principle

8.6.1 Motivation

This is a result we have hinted at before. We even proved a particular *version* of it way back when we studied proof techniques for the first time! (See Example 4.9.2.) The general idea is this:

> If we have too much "stuff" to put into too few "boxes", then some box has a bunch of "stuff".

This is very informal, of course, but it should help you see where it might come in handy.

The Pigeonhole Principple is useful, for instance, when we have a bunch of objects that fall into a certain number of categories. If we know how many objects we have, and how many possible categories there are, then we can guarantee the *existence* of some category that has *at least* some certain number of objects in it.

Example 8.6.1. Here is a canonical example of the principle being applied:

> **Of any 3 people, two must have the same sex.**

Notice that this doesn't say *which* sex is represented at least twice. It just guarantees the *existence* of such a category. To convince yourself of this fact, you could enumerate the possibilities (where M means a male, and F means a female): MMM, MMF, MFF, FFF. In each case, *at least one* of the sexes appears twice (or more).

Here's a logically equivalent version of the above statement:

> If we flip 3 coins, at least two must show the same face.

Here's a similar statement to the ones above:

> If we roll 7 dice, at least two must show the same number.

Are you beginning to see the general idea? Here's one more version of these claims, and a transition to the next part, where we state and prove a generalized version.

> If we have $n + 1$ pieces of paper to stuff into n different drawers, some drawer will end up getting at least 2 pieces of paper.

This is, by the way, the etymological derivation of "pigeonhole": it's a term for the drawers you'd find on an old-fashioned rolltop desk. We'd rather not think about manhandling gentle creatures into tiny boxes!

8.6.2 Statement and Proof

There are two versions of this principle, so we'll state and prove them both. The first version is how we'll be using it in combinatorial problems.

Theorem 8.6.2 (Pigeonhole Principle)**.** *(1) If a set S with $|S| = n$ is partitioned into k disjoint subsets whose union is S, and if $k < n$, then* at least one *of the subsets in the partition has more than one element. Furthermore, that part actually has at least $\lceil \frac{n}{k} \rceil$ elements.*

(That is, if we separate n objects into k piles, there must be one pile with at least $\frac{n}{k}$ objects in it.)

(2) If $x_1, x_2, \ldots, x_n$ are real numbers with the property that $\sum_{i=1}^{n} x_i \geq z$, then there is at least one index i such that $x_i \geq \frac{z}{n}$.

(That is, if we have n real numbers, there must be one number that is at least as large as the average).

Proof. AFSOC $k < n$ and S is partitioned into $S_1, \ldots, S_k$ that also satisfy $|S_i| < \frac{n}{k}$ for every i. Since the sets form a partition of S, we have

$$n = |S| = \sum_{i=1}^{k} |S_i| < \sum_{i=1}^{k} \frac{n}{k} = n$$

so $n < n$. This is a contradiction! ※

AFSOC all of the numbers x_i satisfy $x_i < \frac{z}{n}$. Then,

$$z = \sum_{i=1}^{n} x_i < \sum_{i=1}^{n} \frac{z}{n} = n \cdot \frac{z}{n} = z$$

so $z < z$. This is a contradiction. ※ □

Notice how similar these proofs are! They look identical, algebraically. Indeed, they represent the same underlying idea.

8.6.3 Examples

Let's dive right in and see how to use the Pigeonhole Principle in combinatorics problems. We'll show you how it works with some practice examples. In general, the *hardest* part about using the Pigeonhole Principle is in deciding what the "pigeonholes" actually are!

Example 8.6.3. Of 8 people, there must be two whose birthdays are on the same day of the week this year. Also, of 13 people, there must be two whose birthdays are in the same month.

For the first claim, we can take our "pigeonholes" to be the 7 days of the week. Taking 8 people and partitioning them based on the day of the week their birthday occurs on this year, we find there are 8 objects going into 7 parts. Thus, one part has at least $\frac{8}{7}$ objects in it. Since we are working with *whole* objects, this actually means some part has at least 2 objects in it.

A similar argument applies for the second claim. We just use the 12 months of the years as our "pigeonholes".

Example 8.6.4. In New York City, there are at least 8 people with the exact same number of hairs on their head.

This follows from knowing a couple of facts. First, scientists estimate that there are between 100000 and 150000 hairs on the human head. Let's be conservative and widen that range to 0 to 1 million. Second, New York City has about 8

million people. By defining our "pigeonholes" to be the numbers 0 to 1 million (based on how many hairs are on each person's head), we get the result.

(In fact, this argument might not be necessary. I bet we could walk around the city and find 8 bald people pretty quickly!)

Example 8.6.5. Look back at Section 1.4.4 where we investigated finding a group of mutual friends amongst a larger group. In that problem's solution, we actually used the Pigeonhole Principle! We had 5 objects that were arbitrarily separated into 2 categories. This let us deduce that *some* category had at least 3 of the objects.

Example 8.6.6. Suppose n golfers ($n \geq 2$) compete in a match play tournament, round robin style. How many matches are played? After those matches, must there exist two golfers with the exact same number of wins and losses? If not, can you impose conditions that guarantee that?

Using a counting argument, we find that $\frac{n(n-1)}{2}$ matches are played. (Why? Can you fill in the details? Try it!) However, we can't *guarantee* two people with the same record. For instance, suppose $n = 3$ and that Player 1 lost to both others, Player 2 beat Player 1 but lost to Player 3, and that Player 3 beat both others. This yields records of 0-2 and 1-1 and 2-0, respectively, and we see that none are the same.

Now, if we impose the condition that *no one is undefeated*, then we *can* guarantee two players have the same record. Each player plays $n-1$ matches (one match against everyone except themself). Since no one is undefeated, no one has $n-1$ wins. Thus, the possible number of wins for each player is 0 or 1 or 2 or ... or $n-2$. There are $n-1$ options there. By the Pigeonhole Principle, amongst n players, there must be two of these win counts repeated!

Example 8.6.7. **Proposed claim:** "Amongst any set of m distinct natural numbers, there are at least two such numbers whose sum or difference is a multiple of 10."

Find the smallest value of m such that this claim is valid.

By trying out some small cases, we can see that $m \leq 6$ will *not* work. Even with $m = 6$, we can pick the set of numbers $\{1, 2, 3, 4, 5, 10\}$. Notice that no two of them have a sum/difference that is a multiple of 10. (Note: We aren't allowed to trivially pick the same number twice, like $5 - 5 = 0$ or $5 + 5 = 10$, to get a multiple of 10.)

Might $m = 7$ be the number we are looking for? Let's try to prove it!

Suppose we have an arbitrary set of 7 natural numbers. Let's assign them to the Pigeonhole Boxes that are categorized by their last digit (that is, place each number n in a box based on the smallest positive value of x satisfying $x \equiv n \bmod 10$) as follows: $\{1, 9\}, \{2, 8\}, \{3, 7\}, \{4, 6\}, \{5\}, \{0\}$. That is, we have 6 boxes.

Since we have 7 numbers, then some box has two numbers. That means those numbers either have last digits that sum to 0 modulo 10 (for example 2 and 8 or 5 and 5), or else those numbers have the same last digit so their difference is 0 modulo 10. Either way, we have a sum or difference that is 0 modulo 10, i.e. a multiple of 10.

8.6.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) What are the two versions of the Pigeonhole Principle?

(2) What proof technique did we use to *prove* the Pigeonhole Principle?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) Suppose there are 5 professors in the math department. Every year, 2 are chosen to teach Concepts of Math. How many years can the department go without repeating the same selection of 2 professors? Prove this is optimal by exhibiting such a sequence of that length, as well as invoking the Pigeonhole Principle to show that *any* longer sequence necessarily uses a pairing more than once.

(2) Let $n \in \mathbb{N}$ and consider the set $[2n]$. Suppose we have a set $S \subseteq [2n]$ of size $|S| = n + 1$. Prove that there must be two elements $x, y \in S$ that are *relatively prime*.

(3) Suppose we have a square park with dimensions 1 km $\times$ 1 km. We want to build a golf course on the park, but we only have space for 5 holes. In particular, for safety reasons, we need to consider the distance between the locations of the actual cups (the holes in the ground).

Prove that no matter how we place 5 holes, there must exist two holes that are separated by a distance d that satisfies $d \leq \frac{\sqrt{2}}{2}$ km.

(Note: We *are* allowed to place a hole on the boundary of the park.)

Next, prove that this bound is *optimal*; that is, exhibit a way to place 5

holes on the park grounds (again, the boundary is allowed) such that the distance between any two holes is greater than or equal to $\frac{\sqrt{2}}{2}$ km.

8.7 Inclusion/Exclusion

8.7.1 Motivation

The Principle of Inclusion/Exclusion is a handy result that helps figure out how to *remove* sets from a large one and count the elements leftover. We have already seen this in action in a simple form. If we have $A \subseteq U$, and we want to find $|U - A|$, we can just count $|U|$ and $|A|$ and subtract them. This follows from the Rule of Sum, applied to the partition $\{A, U - A\}$ of the set U.

What happens if we remove two sets from a larger one? What if they overlap somehow? Do we have to account for that? What if we remove three sets? Or four sets? Or n sets? Can we write an expression for the number of elements leftover that holds, in generality? Can we use it to solve counting problems?

Here are some expressions that describe what is going for "small cases". Suppose we have a universal set U and some subsets $A_1, A_2, \ldots, A_n \subseteq U$. We want to count the elements of U that are *outside* of all of the A_i sets. We can do this by writing:

$$\begin{aligned} |U - A_1| &= |U| - |A_1| \\ |U - (A_1 \cup A)2)| &= |U| - |A_1| - |A_2| + |A_1 \cap A_2| \\ |U - (A_1 \cup A_2 \cup A_3)| &= |U| - |A_1| - |A_2| - |A_3| \\ &\quad + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3| \end{aligned}$$

Do you see why these work? Try thinking of an element $x \in U$ and considering how many of the A_i sets it belongs to. Where will this element get counted in the expressions on the left- and right-hand sides? Is it counted the appropriate number of times on both sides? Can you see how to generalize this idea?

It might help to think of these expressions as "guessing" at a correct count and then continually "fixing" to adjust for over/undercounting. For instance, we could derive the last expression above as follows:

> Let's find $|U - (A_1 \cup A_2 \cup A_3)|$. Let's take the number of elements in U and remove the number of elements in the sets A_1, A_2, A_3.
>
> Oh, shucks! What about the elements that belong to *two* of the sets. We have now removed those from our count too many times, so we should add back the number of elements that belong to the intersection of two sets.
>
> Oh, shucks! What about the elements that belong to all *three* sets. We have now added those back in too many times, so we need to remove them again.

You might see now how to generalize these expressions, and prove them, for any number of sets. That is what we will do in the next section.

8.7.2 Statement and Proof

Theorem 8.7.1 (Inclusion/Exclusion)**.** *Suppose we have a universal set U and some subsets $A_1, A_2, \ldots, A_n \subseteq U$. Then,*

$$|U - (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{S \subseteq [n]} (-1)^{|S|} \left| \bigcap_{i \in S} A_i \right| \qquad \text{where} \quad \bigcap_{i \in \varnothing} A_i = U$$

(Try writing out the above expression in the cases where $n = 1$ and $n = 2$ and $n = 3$ to see why they're the same as the ones we wrote in the previous section.

To *prove* this theorem, we will apply a counting in two ways argument. Specifically, we will consider an element $x \in U$ and argue that it is counted the *correct* number of times on both sides of the above equation.

Proof. Let $x \in U$ be arbitrary and fixed. We'll consider two cases.

First, suppose $x \notin A_i$ for every $i \in [n]$. Then the left-hand side counts x exactly once, since x is not an element of the union of the A_i sets. The right-hand side only counts x in the term where $S = \varnothing$, because x is not an element of any of the A_i sets, so it is not an element of any *intersections* thereof. Thus, x is counted exactly once on the right-hand side, as well.

Second, suppose x is an element of one (or more) of the A_i sets. This means x is *not* counted on the left-hand side. To help show that x is also counted *zero* times on the right-hand side, let's define $B \subseteq [n]$ to be the set of indices i for which $x \in A_i$, i.e. $\forall i \in B.\ x \in A_i$. We have a few observations to make:

Consider the terms of the sum that will *not* count x. For any $S \subseteq \mathbb{N}$, if $S \not\subseteq B$, then $x \notin \bigcap_{i \in S} A_i$. (This is because there is some $j \in S$ such that $j \notin B$, but B is all of the indices for which $x \in A_i$.) This means x is counted 0 times in the terms of the sum for which $S \not\subseteq B$.

Next, by a result previously proven, we know that B has an equal number of subsets of odd size as it does of even size. For any such subset $T \subseteq B$, we know $x \in \bigcap_{i \in T} A_i$. Now, if $|T|$ is even, that term will be positive, so x is counted once; if $|T|$ is odd, that term will be negative, so x is removed from the count once. Since there are an equal number of each of these terms, we see that x is accounted for 0 times by these terms, as well.

Overall, we have shown that an arbitrary element is counted the *same* (and *correct*) number of times on both sides of the equation, in any case. □

Sometimes, it happens that all of the intersections of k-many of the A_i sets have the *same size*. This will be true of some of the examples we see in the next section. When that is the case, many of the terms in the expression above can be *combined* since they are identical. Specifically, rather than summing over subsets $S \subseteq [n]$ to account for all possible intersections of the sets A_i, we can sum over the *number* of sets being intersected together, rather than *which* sets are intersected.

Corollary 8.7.2. *Suppose we have a universal set U, and suppose we have some sets $A_1, A_2, \ldots, A_n$ that are all subsets of U. Furthemore, suppose that the intersection of any k of the A_i sets has a fixed size—call it $S(k)$—independent of* which *sets are intersected. Then,*

$$|U - (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{k=0}^{n} (-1)^k \binom{n}{k} S(k)$$

Proof. This follows from the result of Theorem 8.7.1 by combining identical terms. Specifically, we know there are $\binom{n}{k}$ sets $S \subseteq [n]$ that satisfy $|S| = k$. By the assumption of this corollary, all such sets with $|S| = k|$ will yield

$$\left| \bigcap_{i \in S} A_i \right| = S(k)$$

Combining those terms together, and summing over the possible sizes of S, we obtain the above result. □

This will be helpful in some examples we work through below!

8.7.3 Examples

Example 8.7.3. **Bridge hands:**

Bridge deals out 13 cards per hand. How many such hands have at least one card from each suit?

Recall that with poker hands (i.e. 5 cards) this was easy! We just noticed the suit distribution must follow 1112, i.e. some suit appears twice and the others appear once each. (Look back at Example 8.3.6 for the details of this argument.)

With 13 cards, though, it's much harder to write down all of those cases, those partitions of 13 into nonzero terms: (1,1,1,10) and (1,1,2,9) and (1,2,3,7) and so on. There are lots of cases!

Let's use Inclusion/Exclusion to be more efficient.

Let U be the set of all 13 card hands from a standard deck of 52 cards.

Let A_H be the set of 13 card hands that *don't* have any Hearts.

Let A_S be the set of 13 card hands that *don't* have any Spades.

Let A_C be the set of 13 card hands that *don't* have any Clubs.

Let A_D be the set of 13 card hands that *don't* have any Diamonds.

Then we seek an expression for

$$\text{☺} = \left| U - \left(A_H \cup A_S \cup A_C \cup A_D \right) \right|$$

Accounting for all possible intersections, we have

$$\begin{aligned}\text{☺} = {} & |U| - |A_H| - |A_S| - |A_C| - |A_D| \\ & + |A_H \cap A_S| + |A_H \cap A_C| + |A_H \cap A_D| \\ & + |A_S \cap A_C| + |A_S \cap A_D| + |A_C \cap A_D| \\ & - |A_H \cap A_S \cap A_C| - |A_H \cap A_S \cap A_D| \\ & - |A_S \cap A_C \cap A_D| - |A_H \cap A_D \cap A_C| \\ & + |A_H \cap A_S \cap A_C \cap A_D|\end{aligned}$$

Since there are 4 "bad sets", we need to consider all possible ways they can intersect. However, counting these intersections is actually quite convenient because the sizes of the intersection *only* depend on *how many* sets are intersected, not *which* ones they are.

Notice that $|A_H| = |A_S| = |A_C| = |A_D| = \binom{39}{13}$. To have a 13 card hand which *avoids* one set, we just have to select 13 cards from the *other* 39.

Likewise, notice that $|A_H \cap A_S| = \binom{26}{13}$ because we need to avoid 2 suits. This holds for *every* intersection of two of these sets.

Likewise, notice that $|A_H \cap A_S \cap A_D| = \binom{13}{13}$ because we need to avoid 3 suits, so we only have 13 cards to pick from (the 4th suit). This holds for every intersection of three of these sets.

Thus, we have

$$\text{☺} = \binom{52}{13} - \binom{4}{1}\binom{39}{13} + \binom{4}{2}\binom{26}{13} - \binom{4}{3}\binom{13}{13} + \binom{4}{4}\binom{0}{13}$$

total such hands.

(Notice that the last term is 0; how can we have a 13 card hand with no suits represented in it?!)

One Lesson: Notice how we chose to define the sets U and A_i in this example. We wanted to count the number of hands *with* a certain property, so we defined sets of hands that *do not* have that property, and considered how to *remove* their counts from a total.

Example 8.7.4. **Counting surjections:** Count the number of functions $f : [5] \to [3]$. Count the number that are injections. Count the number that are surjections.

Let U be the set of all functions from [5] to [3].

We know $|U| = 3^5$ because we have 3 choices of output for each of the 5 elements in the domain.

There are *no* such functions that are injective. If a function $f : [5] \to A$ is injective, then $|\operatorname{Im}_f([5])| = 5$, but here, the codomain has size 3. Thus, this is not possible.

Now, let's count the surjections!

Let A_1 be the set of all such functions with the property that $1 \notin \text{Im}_f([5])$.

Let A_2 be the set of all such functions with the property that $2 \notin \text{Im}_f([5])$.

Let A_3 be the set of all such functions with the property that $3 \notin \text{Im}_f([5])$.

Then we seek an expression for $N = |U - (A_1 \cup A_2 \cup A_3)|$. We have

$$N = |U| - |A_1| - |A_2| - |A_3| + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3|$$

Remembering that, generally, the number of functions $f : [m] \to [n]$ is n^m (n choices of output for each of m inputs), we have

$$N = 3^5 - \binom{3}{1}2^5 + \binom{3}{2}1^5 - \binom{3}{3}0^5 = 3^5 - 3 \cdot 2^5 + 3 = 243 - 96 + 3 = 150$$

Example 8.7.5. Find the number of natural numbers from 1 to 1000 that are neither perfect squares, cubes, nor fourth powers.

Let $U = [1000]$. For $i \in \{2, 3, 4\}$, let A_i be the set of elements of U that are perfect i-th powers of some natural number. That is, define

$$A_i = \left\{x \in U \mid \exists b \in \mathbb{N}\text{.}\ x = b^i\right\}$$

Then we seek the number $M = |U - (A_2 \cup A_3 \cup A_4)|$.

Notice that $|U| = 1000$.

Notice that the largest square in U is $31^2 = 961$ (since $32^2 = 1024$). Thus, $|A_2| = 31$.

Notice that the largest cube in U is $10^3 = 1000$. Thus, $|A_3| = 10$.

Notice that the largest fourth power in U is $5^4 = 625$ (since $6^4 = 1296$). Thus, $|A_4| = 5$.

Considering intersections, notice that, for instance, $A_2 \cap A_3$ is the set of sixth powers since $\text{LCM}(2, 3) = 6$. (LCM is the *Least Common Multiple*.)

Notice that the largest sixth power in U is $3^6 = 729$ (since $4^6 = 4096$). Thus, $|A_2 \cap A_3| = 3$.

We already found the largest fourth power in U to be 5^4, so $|A_2 \cap A_4| = |A_4| = 5$.

Notice that the largest 12th power in U is $1^{12} = 1$ (since $2^{12} = 4096$). Thus, $|A_3 \cap A_4| = 1$.

This also tells us that $|A_2 \cap A_3 \cap A_4| = |A_3 \cap A_4| = 1$.

Putting this all together, we find

$$N = 1000 - 31 - 10 - 5 + 3 + 5 + 1 - 1 = 962$$

8.7.4 Questions & Exercises

Remind Yourself

Answering the following questions briefly, either out loud or in writing. These are all based on the section you just read, so if you can't recall a specific definition or concept or example, go back and reread that part. Making sure you can confidently answer these before moving on will help your understanding and memory!

(1) When is the Principle of Inclusion/Exclusion applicable?

(2) What strategy did we use to prove the Principle of Inclusion/Exclusion?

(3) Why did we require $A_i \subseteq U$ for each i? Do you think the result still holds if these conditions are not satisfied?

Try It

Try answering the following short-answer questions. They require you to actually write something down, or describe something out loud (to a friend/classmate, perhaps). The goal is to get you to practice working with new concepts, definitions, and notation. They are meant to be easy, though; making sure you can work through them will help you!

(1) How many natural numbers less than 100 are not multiples of 2 or 5?

(2) How many natural numbers less than 1000 are not perfect powers of 2 or 3 or 5?

(3) How many lattice paths go from $(0, 0)$ to $(10, 10)$ *without* going through $(3, 3)$?

(4) How many lattice paths go from $(0, 0)$ to $(10, 10)$ *without* going through either $(3, 3)$ or $(6, 8)$?

(5) How many functions $f : [6] \to [3]$ are surjections?

8.8 Summary

We have now developed several basic counting techniques and developed them into even more advanced techniques. We started by simply discussing the Rules of Sum and Product, which were based on results from the previous chapter about the cardinality of finite sets. We were able to use these to develop some fundamental counting objects and describe how to count them. This included the vastly useful **binomial coefficients**. We derived the formula for binomial coefficients for ourselves, implement a counting strategy. Then, we applied these principles to plenty of examples, to give ourselves practice with working through the nuances of counting arguments: sometimes there are many cases involved,

sometimes we have to be clever about applying the Rule of Product, sometimes we need to be worried about an over/undercount. On that note, we discussed how to take a proposed argument and *demonstrate* that it is incorrect.

The proof technique of *counting in two ways* is incredibly important, and you will see it appear in plenty of other mathematical areas. We saw some instructive examples—which were useful theorems in their own right—and have posed many problems of this kind in the exercises to give you sufficient practice. We used the counting in two ways technique to later prove some further results and techniques, including the Binomial Theorem, the formula for selections with repetition.

We briefly discussed some more advanced counting techniques, the Pigeonhole Principle and Inclusion/Exclusion. These are considered to be more advanced partly because it can be difficult to see *when* and *how* to apply them. By working through some illustrative examples, we hope we have given you a better intuition for how these techniques can be useful, so that you will see when to use them in problem-solving.

8.9 Chapter Exercises

These problems incorporate all of the material covered in this chapter, as well as any previous material we've seen, and possibly some assumed mathematical knowledge. We don't expect you to work through **all** of these, of course, but the more you work on, the more you will learn! Remember that you can't truly *learn* mathematics without *doing* mathematics. Get your hands dirty working on a problem. Read a few statements and walk around thinking about them. Try to write a proof and show it to a friend, and see if they're convinced. Keep practicing your ability to take your thoughts and *write* them out in a clear, precise, and logical way. Write a proof and then edit it, to make it better. Most of all, just keep *doing* mathematics!

Short-answer problems, that only require an explanation or stated answer without a rigorous *proof*, have been marked with a ►.

Particularly challenging problems have been marked with a ★.

Problem 8.9.1. In this problem, you will *prove* the **Rule of Product** (see Theorem 8.2.10).

Prove, by induction on n, that the Cartesian product of n finite sets has size equal to the product of the sizes of those sets.

Problem 8.9.2. Let $n \in \mathbb{N}$ (with $n \geq 3$) and let S be the set of all binary strings of length n.

Each of the following expressions is the size of some subset of S. For each one, identify such a subset and explain why it works.

For example, if I were presented with

$$\binom{n}{3} + \binom{n}{4} + \binom{n}{5}$$

I would say,

> Let $S_1 \subseteq S$ be the set of all strings with either 3 or 4 or 5 positions that are 0s. We can partition this set into the set of strings with exactly k positions that are 0s, for each $k = 3, 4, 5$. In each case, we can find the size of that part by selecting k of the n total positions to be 0s, and fixing the rest to be 1s. By ROS, then, we find that $|S_1|$ is the sum above.

(a) 2^{n-2}

(b) $2^n - \binom{n}{n} - \binom{n}{n-1} - \binom{n}{n-2} - \binom{n}{n-3}$

(c) $\binom{n}{2} - \binom{n-1}{1}$

(d) $\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k}$

Problem 8.9.3. A student organization holds meetings every week, with one chosen leader and two assistants to run the meeting efficiently. If there are 14 weeks in the semester, how many students must be in the organization to guarantee that they can have a different set of leaders and assistants at every meeting?

Problem 8.9.4. Say we have 50 pieces of Halloween candy to distribute amongst 4 different children. How many ways can we do this? What if all of the pieces of candy are identical? What if there are 10 pieces each of 5 different kinds?

Problem 8.9.5. Let U be the set of all 5-card hands, as dealt from a standard deck of cards, that have exactly two Kings and exactly one Heart. Find $|U|$.

Problem 8.9.6. For each of the following conditions, consider drawing a 7-card hand from a standard deck of cards.

(a) How many 7 card hands are there?

(b) How many 7 card hands have no cards ranked above 8? (Note: A is the highest rank.)

(c) How many 7 card hands have exactly two Kings?

(d) How many 7 card hands include contain exactly one pair? (That is, one pair and five other cards of different ranks.)

(e) How many 7 card hands have at least 3 ♡s?

Problem 8.9.7. Find the number of ordered arrangements of 5 *distinct* digits from $\{0, 1, 2, \ldots, 9\}$. Then, find the number of such arrangements that do *not* place 5 and 6 adjacent to each other.

Problem 8.9.8. Let $T_{5,4}$ be the set of all 5-tuples drawn from the set [4].

(For example, $(1, 4, 4, 1, 2) \in T_{5,4}$.)

(a) What is $|T_{5,4}|$?

(b) How many elements of $T_{5,4}$ have no odd numbers?

(c) How many elements of $T_{5,4}$ have no repeated numbers?

(d) How many elements of $T_{5,4}$ have exactly 2 distinct numbers?

(For example, $(1, 2, 2, 1, 2)$ should be counted but $(1, 1, 1, 1, 1)$ should not, and neither should $(1, 2, 3, 3, 3)$.)

(e) How many elements of $T_{5,4}$ have no adjacent identical numbers?

(For example, $(1, 3, 1, 3, 4)$ should be counted but $(2, 3, 1, 1, 3)$ should not, and neither should $(1, 1, 1, 4, 3)$.)

Problem 8.9.9. For each of the following properties, find the number of ways to roll 5 *distinguishable* dice so that the property holds. (Don't consider the combination of properties; each one is separate).

(a) No even numbers appear on the faces.

(b) Exactly 2 even numbers appear on the faces.

(c) The sum of all the faces is odd.

(d) The numbers on the faces form a "full house". (That is, there are exactly three of one number and exactly two of another.)

(e) The numbers on the faces form a "straight".

Problem 8.9.10. How many anagrams are there of the word MILLIMETER? How many such anagrams have the two Ms adjacent? How many such anagrams have the Ms *non*-adjacent?

Problem 8.9.11. How many natural numbers between 1 and 1000 (inclusive) have the property that none of the digits are even? How many have the property that no digits are repeated? How many have the property that the sum of the digits is even?

(Be careful: Remember that a string like 0011 is really the number 11.)

Problem 8.9.12. Consider drawing two cards **in order** from the top of a deck of cards. How many outcomes are such that the first card is an Ace and the second card is a Heart?

Problem 8.9.13. How many 15 card hands have at least one card from each rank?

Problem 8.9.14. Prove the Binomial Theorem (see Theorem 8.4.8) by **induction** on n.

Problem 8.9.15. Prove that

$$\binom{n}{k}2^k = \sum_{i=0}^{k} \binom{n}{i}\binom{n-i}{k-i}$$

Problem 8.9.16. Let $a, b, k \in \mathbb{N}$ with $a + b \geq k$. Prove that

$$\binom{a+b}{k} = \sum_{i=0}^{k} \binom{a}{i}\binom{b}{k-i}$$

Problem 8.9.17. Three men walk into the bathroom and find seven urinals in a row on the wall. In how many ways can the men arrange themselves so that they don't violate the "bro code"? (That is, they must make sure no two adjacent urinals are occupied.)

Problem 8.9.18. Let $n \in \mathbb{N}$ be given. Prove the following identities by *counting in two ways* arguments.

(**Hint:** It's likely that you can use the same "story" or formulation in all parts; that is, try slightly modifying your argument from (a) to prove (b) and (c).)

(a) $\displaystyle\sum_{i=1}^{n}(i-1) = \binom{n}{2}$

(b) $\displaystyle\sum_{i=1}^{n}(i-1)(n-i) = \binom{n}{3}$

(c) $\displaystyle\sum_{i=1}^{n}\binom{i-1}{2}\binom{n-i}{2} = \binom{n}{5}$

Problem 8.9.19. Prove that

$$\sum_{i=0}^{n}\binom{r+i}{i} = \binom{r+n+1}{n}$$

by a counting in two ways argument.

Problem 8.9.20. Prove that

$$\binom{n}{k} - \binom{n-2}{k} = 2\binom{n-2}{k-1} + \binom{n-2}{k-2}$$

by a counting in two ways argument. Use the exact form given; do not simplify algebraically.

Problem 8.9.21. Prove that

$$\binom{n}{k} - \binom{n-2}{k} = \binom{n-1}{k-1} + \binom{n-2}{k-1}$$

by a counting in two ways argument. Use the exact form given; do not simplify algebraically.

Problem 8.9.22. Prove that

$$\binom{n}{k} - \binom{n-3}{k} = \binom{n-1}{k-1} + \binom{n-2}{k-1} + \binom{n-3}{k-1}$$

by a counting in two ways argument. Use the exact form given; do not simplify algebraically.

Problem 8.9.23. Prove that

$$4^n = \sum_{k=0}^{n} \binom{n}{k} 3^k$$

by a counting in two ways argument.

Problem 8.9.24. Let $p \in \mathbb{N}$ be prime. Let $k \in \mathbb{N}$ be given with $1 \leq k < p$. Prove that $\binom{p}{k}$ is divisible by p.

Problem 8.9.25. Let $p \in \mathbb{N}$ be prime. Use the previous Problem 8.9.24 to prove that

$$\forall x, y \in \mathbb{Z}\textbf{.}\ (x+y)^p \equiv x^p + y^p \mod p$$

(Look back at Problem 6.7.22 where we investigated this before. You just proved it in generality!)

Problem 8.9.26. Let $p \in \mathbb{N}$ be prime, and let $a \in \mathbb{Z}$. Use the result of Problem 8.9.24 and the Binomial Theorem to prove that

$$a^p \equiv a \mod p$$

This result is known as **Fermat's Little Theorem**.

Problem 8.9.27. Prove the Summation Identity (see Theorem 8.4.5) by a counting in two ways argument that considers *lattice paths*. Specifically, we suggest partitioning the set of lattice paths from $(0,0)$ to $(k+1, n-k)$ based on where the first Rightwards step occurs.

Problem 8.9.28. In this problem, you will prove the following summation formula that you've proved by induction before!

$$\forall n \in \mathbb{N}\textbf{.} \quad \sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$$

(a) Let $k \in \mathbb{N}$ be given. Prove the following equality by a *counting in two ways* argument:

$$\forall k \in \mathbb{N}\textbf{.} \quad k^3 = 6\binom{k}{3} + 6\binom{k}{2} + \binom{k}{1}$$

(Hint: Consider counting words of length 3 from an alphabet of k letters.)

(b) Use the **Summation Identity** and the equality you just proved in (a) to prove the claim given above in the problem statement!

Bonus Can you generalize this method to find a formula for $\sum k^4$?

Problem 8.9.29. Let $n \in \mathbb{N}$ be given. How many lattice paths go from $(0,0)$ to $(3n,3n)$ without going through either of (n,n) or $(2n,2n)$?

Problem 8.9.30. Let $n \in \mathbb{N}$ be given. Suppose we have n CMU students and n Pitt students. (Assume, of course, that nobody attends both schools, so these two sets of n students are disjoint.)

(a) How many ways can we split these $2n$ students into n pairs? (Note: There should be no ordering to the pairs, nor the people within the pairs.)

(b) How many ways can we split these $2n$ students into n pairs, where each pair must contain one CMU student and one Pitt student? (Again, there is no order amongst or within the pairs.)

Problem 8.9.31. Let $n \in \mathbb{N}$, and let $S \subseteq \mathbb{N}$ be of size $|S| = n+1$. Prove that $\exists x, y \in S$ such that $x \neq y$ and $x - y$ is a multiple of n.

Problem 8.9.32. Consider the set $[22]$. Let $S \subseteq [22]$ be given such that $|S| = 7$. Here, you will prove that there must exist two disjoint, non-empty subsets $X, Y \subseteq S$ whose elements have the same *sum*.

1. How many non-empty subsets of S are there?
2. Let $T \subseteq S$ be given. What is the minimum possible value of the sum of the elements of T? What is the maximum possible value?
3. Use (a) and (b) to deduce that there are two sets $X, Y \subseteq S$ whose elements have the same sum.
4. Explain, further, that you can make X and Y be *disjoint*.

Problem 8.9.33. Consider an equilateral triangle with side length 1 cm. Suppose 10 points have been placed inside the triangle (and not on the boundary). Prove that there must be two points separated by a distance d that is less than $\frac{1}{3}$ cm.

Problem 8.9.34. Let $n \in \mathbb{N}$ be given. Prove the following identity by a *counting in two ways* argument:

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$$

Problem 8.9.35. Let $n \in \mathbb{N}$ be given. Consider the following identity:

$$4^n = \sum_{k=0}^{n} \binom{n}{k} 2^n$$

Deduce it from the Binomial Theorem. Then, prove it by a *counting in two ways* argument.

Problem 8.9.36. Let $n, k \in \mathbb{N}$ be given. Consider Equation $\star$:

$$\sum_{i \in [k]} x_i = x_1 + x_2 + \cdots + x_k = n$$

In this problem, we will discuss *solutions* to $\star$, where a solution is an assignment of values for $x_1, x_2, \ldots, x_k$ such that their sum is n and such that each one satisfies $x_i \in \mathbb{N} \cup \{0\}$.

(a) How many solutions to $\star$ exist?

(b) How many solutions to $\star$ also satisfy $x_1 \geq 3$?

(c) How many solutions to $\star$ also satisfy $\forall i \in [k]\,.\; x_i \geq 2$?

(d) How many solutions to $\star$ also satisfy $x_1 \leq 4$?

(e) Consider the following modification to $\star$:

$$x_1 + x_2 + \cdots + x_k \leq n$$

How many solutions to this *inequality* exist? (Again, a solution requires $x_i \in \mathbb{N} \cup \{0\}$.)

Problem 8.9.37. Suppose we have 10 pirates who need to divvy up 100 pieces of gold.

Suppose Captains Redbeard and Blackbeard are among the 10 pirates.

(a) How many ways are there to divvy the gold so that Redbeard gets at least 5 pieces of gold but Blackbeard gets at most 5 pieces?

(b) How many ways are there to divvy the gold so that Redbeard gets at least 10 pieces of gold but Blackbeard gets somewhere between 5 and 15 (inclusive) pieces of gold?

(c) How many ways are there to divvy the gold so that Redbeard gets somewhere between 0 and 10 (inclusive) pieces of gold and Blackbeard gets somewhere between 10 and 20 (inclusive) pieces of gold?

(**Hint:** Use Inclusion/Exclusion.)

8.10 Lookahead

There isn't much to look ahead for, at this point! At least, there isn't anything else in *this* book. We hope this has only served to whet your appetite for mathematical knowledge and problem-solving. Think about where we started: we were doing nothing more than posing fun puzzles and trying to solve them by applying our current knowledge and logical techniques. In reality, that's all we are doing now, as well! It's just that we have developed so much mathematical

terminology and so many results and so many problem-solving skills that we are able to approach and discuss far more advanced problems. Did you think, when you started reading this book, that you would be solving problems like this? Do you feel like you have a better understanding, an appreciation, of what mathematicians work on and how they approach the world? We hope so! ☺

We have also developed several essential skills that are applicable in life outside of mathematics, as well. It's likely that you won't encounter formal *symbolic* logic in your daily life, but you will certainly have to process complicated statements with conjunctions and disjunctions and conditionals. We do this on a daily basis, as human beings who converse with one another and convey complex thought processes. By working through some of the foundational aspects of formal logic, we are all better now at analyzing complicated statements and assessing their truth, as well as being able to write down or otherwise share our own thoughts.

Likewise, you might not face formal statements of combinatorial identities in your daily life, but our work with counting in two ways proofs will help your anlytical thinking skills. We had to sometimes be creative about how to develop a "story" to describe a set of elements to be counted in two ways. This required some creativity and ingenuity, and exercising those brain muscles can only be helpful. Furthermore, the practice of reading a proposed "proof" and analyzing whether or not it is, in fact, an over/undercount has made us better at understanding and critiquing the arguments of others. Surely, this is something you do on a daily basis, but perhaps not in mathematical terms.

Overall, we have developed an ability to think in a mathematical way. We have learned: how to read and understand a problem; how to approach a problem from several angles, and be willing to pursue potential dead ends to further our understanding; how to identify common structures that underlie different problems and exploit these similarities with certain techniques; and ultimately, how to take everything we have figured out about a problem and formulate our ideas into a written, presentable argument to be read by others. This entire process will make us all not only better problem-solvers, but better *communicators*. In a rapidly-changing world where communication is more and more important (as it becomes easier and easier to connect to other people quickly), the ability to share our thoughts effectively, correctly, and clearly is an essential life skill.

But by all means, don't let this be the end of our mathematical journey! Go forth and prosper, and spread your knowledge and joy of mathematics. Work on these problems and more with other people. Seek out the areas of mathematics that inspire you. See if you can use these concepts to solve a real-world problem you're facing. Most of all, just get out there and **do mathematics**.

Appendix A

Definitions and Theorems

A.1 Sets

A.1.1 Standard Sets

- The **natural numbers** are
$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$$
Note: $0 \notin \mathbb{N}$.
- For every $n \in \mathbb{N}$, the set $[n]$ ("**brackets** n") is defined by
$$[n] = \{x \in \mathbb{N} \mid 1 \leq x \leq n\} = \{1, 2, 3, \dots, n\}$$
- The **integers** are
$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$
- The **rational numbers** are
$$\mathbb{Q} = \left\{x \in \mathbb{R} \mid \exists a, b \in \mathbb{Z}\text{.}\ b \neq 0 \text{ and } \frac{a}{b} = x\right\}$$
- The **real numbers** are denoted by $\mathbb{R}$. Every real number is either **rational** or **irrational**.
- The **empty set** is the set that has no elements. We write it as $\varnothing$ or $\{\ \}$.

A.1.2 Set-Builder Notation

- If U is a set and $P(x)$ is some **property** that either does or does not hold for any given x, then we can always define a new set by writing
$$S = \{x \in U \mid P(x) \text{ holds}\}$$
- This is called **set-builder notation**. It is essential to identify the **universal set** U and the **property** $P(x)$.

A.1.3 Elements and Subsets

- To say "x is an element of the set S" we write

  $$x \in S$$

  To say "x is not an element of the set S" we write

  $$x \notin S$$

- To say "S is a subset of T" we write

  $$S \subseteq T$$

  This is defined by the conditional statement "Every element of S is also an element of T". This can be expressed as

  $$\forall x \in U \,.\, x \in S \implies x \in T$$

  That is, for every element x of the universal set (supposing $S, T \subseteq U$), whenever $x \in S$, we also know that $x \in T$.

- To prove that a set is a subset of another set, like $S \subseteq T$, we need to do something like this:

  > Let $x \in S$ be arbitrary and fixed.
  >
  > ... blah blah blah ...
  >
  > Therefore, $x \in T$, as well.
  >
  > This shows $S \subseteq T$.

- To say "S is a proper subset of T" we write

  $$S \subset T$$

  This means $S \subseteq T$ and $S \neq T$.

- It is true that $\varnothing \subseteq S$, for any set S.

- It is true that $S \subseteq S$, for any set S.

A.1.4 Power Set

- Let S be a set. The **power set of** S is denoted by $\mathcal{P}(S)$ and is defined by

  $$\mathcal{P}(S) = \{A \mid A \subseteq S\}$$

  That is, $\mathcal{P}(S)$ is the **set of all subsets of** S.

- It is true that $\varnothing \in \mathcal{P}(S)$ and $S \in \mathcal{P}(S)$, for any set S.

A.1.5 Set Equality

- To say "S and T are **equal** sets", we write $S = T$. This is defined by

$$S = T \text{ if and only if } S \subseteq T \text{ and } T \subseteq S$$

- To **prove** two sets are equal, like $S = T$, we need to do something like this:

> First, we will prove $S \subseteq T$. Let $x \in S$ be arbitrary and fixed.
>
> ... blah blah blah ...
>
> Therefore, $x \in T$. This shows $S \subseteq T$.
>
> Second, we will prove $T \subseteq S$. Let $y \in T$ be arbitrary and fixed.
>
> ... blah blah blah ...
>
> Therefore, $x \in S$. This shows $T \subseteq S$.
>
> Therefore, $S = T$.

  This is known as a **double-containment argument**.

A.1.6 Set Operations

Suppose S, T, U are sets and $S \subseteq U$ and $T \subseteq U$.

- The **union** of two sets is defined by

$$S \cup T = \{x \in U \mid x \in S \text{ or } x \in T\}$$

  It is the set of all elements that belong to **at least one** of the two sets, S and T.

- The **intersection** of two sets is defined by

$$S \cap T = \{x \in U \mid x \in S \text{ and } x \in T\}$$

  It is the set of all elements that belong to **both** sets, S and T.

- The **difference** of two sets is defined by

$$S - T = \{x \in U \mid x \in S \text{ and } x \notin T\}$$

  It is the set of all elements of S that are not elements of T.

- The **complement** of a set is defined by

$$\overline{S} = \{x \in U \mid x \notin S\} = U - S$$

  It is the set of all elements of the universal set that are not elements of S.

- The **Cartesian product** of two sets is defined by

$$S \times T = \{(x, y) \mid x \in S \text{ and } y \in T\}$$

  It is the set of all **ordered pairs**, where the first coordinate is an element of S and the second coordinate is an element of T.

A.1.7 Indexed Set Operations

Suppose I is an index set and U is a universal set, and we have defined (for every $i \in I$) some sets $A_i \subseteq U$.

- The **indexed union** of all of the A_i sets is defined by

$$\bigcup_{i \in I} A_i = \{x \in U \mid \exists k \in I \,.\, x \in A_k\}$$

  It is the set of all elements x in the universal set such that x is an element of **at least one** of the indexed sets in the union.

- The **indexed intersection** of all of the A_i sets is defined by

$$\bigcap_{i \in I} A_i = \{x \in U \mid \forall i \in I \,.\, x \in A_i\}$$

  It is the set of all elements x in the universal set such that x is an element of **all** of the indexed sets in the intersection.

A.1.8 Partition

- Let S be a set. A **partition** of S is a collection of sets that are pairwise disjoint and whose union is S. That is, a partition is formed by an index set I and *non-empty* sets S_i (defined for every $i \in I$) that satisfy:
  - $\forall i \in I \,.\, S_i \neq \varnothing$
  - $\forall i \in I \,.\, S_i \subseteq S$
  - $\forall i, j \in I \,.\, i \neq j \implies S_i \cap S_j = \varnothing$
  - $\bigcup_{i \in I} S_i = S$

A.2 Logic

A.2.1 Statements and Propositions

- True and False are the only two truth values we consider.
- A **mathematical statement** (or **logical statement**) is a grammatically-correct sentence that has **exactly** one truth value.
- A **variable proposition** is a grammatically-correct sentence that involves one or more variables, such that it acquires exactly one truth value whenever a value(s) for the variable(s) is assigned.
- When we define a statement or proposition, we assign it a letter name, indicate any dependence on variables (as well as assigning them letters), and enclose the actual statement/proposition within quotation marks. Here are two **good examples**:

  Define P to be "Every real number x satisfies $x^2 \geq 0$".

  Define $Q(x, y)$ to be " $xy \leq \left(\frac{x+y}{2}\right)^2$ ", for every $x, y \in \mathbb{R}$.

- The **Law of the Excluded Middle** is our assumption that every statement is either True or False. It states that, when we have a statement P, we are guaranteed that either P is True or P is False, and **only one** of those cases holds.

A.2.2 Quantifiers

- To say "for every" or "for all" we use the **universal quantifier** $\forall$

  "$\forall x \in S\,.\; P(x)$" says that "For every element $x \in S$, the property $P(x)$ holds true".

- To say "there exists" or "there is at least one" we use the **existential quantifier** $\exists$

  "$\exists x \in S\,.\; P(x)$" says that "There exists an element $x \in S$ with the property $P(x)$".

- We use the "." dot to separate parts of a quantified statement.
- When reading a quantified statement out loud, we say "such that" **only** after a $\exists$ quantifier.
- We use "!" to indicate that existence is **unique**; that is, the claim "$\exists\,!\, x \in S\,.\; P(x)$" says that "There exists an element $x \in S$ with property $P(x)$, and there is *exactly* one such x".

A.2.3 Connectives

Suppose P and Q are mathematical statements. They may be composed of variable propositions with quantifiers in front.

- To say "P and Q" we write
$$P \wedge Q$$
This has the truth value True if and only if **both** P and Q are True.

- To say "P or Q" we write
$$P \vee Q$$
This has the truth value True if and only if **at least one** of the statements, P and Q, are true. (This is the **inclusive** or, so it's allowable that both P and Q are true.)

- To say "If P then Q" we write
$$P \implies Q$$
This has the truth value True if and only if, whenever P holds, Q also holds.

  Notice that $P \implies Q$ is, itself, a logical statement. It has a truth value, True or False. It makes **no claim** about the truth values of the constituent statements, P and Q.

  We call this a **conditional statement**; we say P is the **hypothesis** and Q is the **conclusion**.

  Notice that $P \implies Q$ is True when P is False. This is because it is an "If ... then ..." statement; it makes **no claim** about the situations where P is False, so we cannot declare the conditional statement to be False so it must be True (by the Law of the Excluded Middle).

- An equivalent way to write $P \implies Q$ is
$$\neg P \vee Q$$

- The **contrapositive** of a condtional statement $P \implies Q$ is
$$\neg Q \implies \neg P$$
It is guaranteed to have the same truth value of $P \implies Q$. That is,
$$(P \implies Q) \iff (\neg Q \implies \neg P)$$

- The **converse** of a condtional statement $P \implies Q$ is
$$Q \implies P$$

It is **not** guaranteed to have the same truth value as $P \implies Q$. There are statements P, Q such that $P \implies Q$ holds and the converse holds, and there are statements P, Q such that $P \implies Q$ holds and the converse fails.

- To say "P and Q are **logically equivalent**" we write

$$P \iff Q$$

and we read this aloud as "P if and only if Q".

We can also write this as

$$(P \implies Q) \wedge (Q \implies P)$$

This means that P and Q **have the same truth value**, whatever that happens to be.

A.2.4 Logical Negation

- We use "$\neg$" to indicate the **logical negation** of a statement.
- The statement $\neg P$ has the **opposite truth value** from the statement P.
- **Negating a $\forall$ claim:**

$$\neg\big(\forall x \in S\text{.}\ P(x)\big) \iff \exists x \in S\text{.}\ \neg P(x)$$

- **Negating a $\exists$ claim:**

$$\neg\big(\exists x \in S\text{.}\ P(x)\big) \iff \forall x \in S\text{.}\ \neg P(x)$$

- **Negating a $\vee$ claim:**

$$\neg\big(P \vee Q\big) \iff \neg P \wedge \neg Q$$

This is one of **DeMorgan's Laws for Logic**.

- **Negating a $\wedge$ claim:**

$$\neg\big(P \wedge Q\big) \iff \neg P \vee \neg Q$$

This is one of **DeMorgan's Laws for Logic**.

- **Negating a $\implies$ claim:**

$$\neg\big(P \implies Q\big) \iff \neg(\neg P \vee Q) \iff P \wedge \neg Q$$

- **Negating a $\iff$ claim:**

$$\neg\big(P \iff Q\big) \iff \neg\left[(P \implies Q) \wedge (Q \implies P)\right] \iff (P \wedge \neg Q) \vee (Q \wedge \neg P)$$

- Using these facts, we can negate **any** mathematical statement, because a statement is just composed of quantifiers and connectives and variable propositions.

We can read the statement left to right and negate each part.

A.2.5 Proof Strategies

We use the phrase AFSOC to mean "assume for sake of contradiction".

- **Proving a $\exists$ claim:** $\exists x \in S\,.\ P(x)$

  Direct proof:

  Define a specific example, $y =$ _____.

  Prove that $y \in S$.

  Prove that $P(y)$ holds true.

  Indirect proof:

  AFSOC that for every $y \in S$, $\neg P(y)$ holds.

  Find a contradiction.

- **Proving a $\forall$ claim:** $\forall x \in S\,.\ P(x)$

  Direct proof:

  Let $y \in S$ be arbitrary and fixed.

  Prove that $P(y)$ holds true.

  Indirect proof:

  AFSOC that $\exists y \in S$ such that $\neg P(y)$ holds.

  Find a contradiction.

- **Proving a $\vee$ claim:** $P \vee Q$

  Direct proof:

  Prove that P holds true, or else prove that Q holds true.

  Indirect proof 1:

  Suppose that $\neg P$ holds. Prove that Q holds.

  Indirect proof 2:

  AFSOC that $\neg P \wedge \neg Q$ holds. Find a contradiction.

- **Proving a $\wedge$ claim:** $P \wedge Q$

  Direct proof:

  Prove that P holds. Prove that Q holds.

  Indirect proof:

AFSOC that $\neg P \vee \neg Q$ holds.

Consider the first case, where $\neg P$ holds. Find a contradiction.

Consider the second case, where $\neg Q$ holds. Find a contradiction.

- **Proving a $\implies$ claim:** $P \implies Q$

  Direct proof:

  Suppose P holds. Prove that Q holds.

  Contrapositive proof:

  Suppose that $\neg Q$ holds. Prove that $\neg P$ holds.

  Indirect proof:

  AFSOC that P holds and suppose that Q fails. Find a contradiction.

- **Proving a $\iff$ claim:** $P \iff Q$

  Direct proof:

  Prove that $P \implies Q$ (using one of the methods above).

  Prove that $Q \implies P$ (using one of the methods above).

  Indirect proof:

  AFSOC that $\neg(P \implies Q) \vee \neg(Q \implies P)$.

  Consider the first case, where $P \wedge \neg Q$ holds. Find a contradiction.

  Consider the second case, where $Q \wedge \neg P$ holds. Find a contradiction.

A.3 Induction

A.3.1 Principle of Specific Mathematical Induction

- **Theorem:** Suppose that $P(n)$ is a variable proposition that is defined for all $n \in \mathbb{N}$.

  Suppose that $P(1)$ holds.

  Suppose that $\forall k \in \mathbb{N}.\ P(k) \implies P(k+1)$.

  Then $\forall n \in \mathbb{N}.\ P(n)$.

- **Proving a claim by induction:** Suppose we have a variable proposition $P(n)$ that is defined for all $n \in \mathbb{N}$, and we want to prove $P(n)$ holds for every $n \in \mathbb{N}$.

  Base Case: Prove that $P(1)$ holds.

  Induction Hypothesis: Suppose that k is an arbitrary and fixed natural number, and suppose that $P(k)$ holds.

  Induction Step: Prove that $P(k+1)$ holds.

  Conclusion: By induction, $\forall n \in \mathbb{N}.\ P(n)$.

A.3.2 Principle of Strong Mathematical Induction

- **Theorem:** Suppose that $P(n)$ is a variable proposition that is defined for all $n \in \mathbb{N}$.

  Suppose that $P(1)$ holds.

  Suppose that $\forall k \in \mathbb{N}.\ [P(1) \wedge P(2) \wedge \cdots \wedge P(k)] \implies P(k+1)$.

  Then $\forall n \in \mathbb{N}.\ P(n)$.

- **Proving a claim by strong induction:** Suppose we have a variable proposition $P(n)$ that is defined for all $n \in \mathbb{N}$, and we want to prove $P(n)$ holds for every $n \in \mathbb{N}$.

  Base Case(s): Prove that $P(1)$ holds.

  (Depending on what happens in the Induction Step, you might need more than one Base Case.)

  Induction Hypothesis: Suppose that k is an arbitrary and fixed natural number that satisfies some inequality ($k \geq ___$, depending on what happens in the Induction Step), and suppose that $P(1) \wedge \cdots \wedge P(k)$ holds.

  Induction Step: Prove that $P(k+1)$ holds.

  Conclusion: By induction, $\forall n \in \mathbb{N}.\ P(n)$.

A.3.3 "Minimal Criminal" Argument

- The second condition in the hypothesis of the Principle of Induction is a conditional statement, so we can prove it by contrapositive. The contrapositive says

$$\neg P(k) \implies \neg P(1) \vee \neg P(2) \vee \cdots \vee \neg P(k-1)$$

  which is to say, "If the proposition fails for some particular value k, then we can find some *prior* instance of the proposition (from 1 to $k-1$) that also fails.

- **Proving a claim by a "minimal criminal" argument:** Suppose we have a variable proposition $P(n)$ that is defined for all $n \in \mathbb{N}$, and we want to prove $P(n)$ holds for every $n \in \mathbb{N}$.

  Base Case(s): Prove that $P(1)$ holds.

  (Depending on what happens in the Induction Step, you might need more than one Base Case.)

  Induction Hypothesis: Suppose that k is an arbitrary and fixed natural number that satisfies some inequality ($k \geq ___$, depending on what happens in the Induction Step), and suppose that $\neg P(k)$ holds; that is, suppose $P(k)$ fails to hold.

  Induction Step: Prove that $\neg P(1) \vee \neg P(1) \vee \neg P(2) \vee \cdots \vee \neg P(k-1)$; that is, show that the proposition fails to hold at some prior instance, as well.

  Conclusion: By induction, $\forall n \in \mathbb{N}\,.\; P(n)$.

A.4 Relations

- Let A, B be sets. A **relation** between A and B is a set of *ordered pairs* $R \subseteq A \times B$.

  Given elements $a \in A$ and $b \in B$, we say a and b are **related** if and only if $(a, b) \in R$.

  The set A is called the **domain** and the set B is called the **codomain**.

  The set R is called the **relation set**.

  We say R is a relation **between** A **and** B.

  When $A = B$, we say R is a relation **on the set** A.

A.4.1 Properties of Relations

Let A be a set and let R be a relation on A, i.e. $R \subseteq A \times A$.

(Note: These properties *only* apply in this case, and not to a relation between two *different* sets A and B.)

- We say R is **reflexive** if

  $$\forall x \in A \bullet (x, x) \in R$$

  (i.e. every element is related to itself).

- We say R is **symmetric** if

  $$\forall x, y \in A \bullet (x, y) \in R \implies (y, x) \in R$$

  (i.e. the order of the comparison doesn't matter).

- We say R is **transitive** if

  $$\forall x, y, z \in A \bullet [(x, y) \in R \land (y, z) \in R] \implies (x, z) \in R$$

  (i.e. the relation always "transitions through a middle-man")

- We say R is **anti-symmetric** if

  $$\forall x, y \in A \bullet [(x, y) \in R \land (y, x) \in R] \implies x = y$$

  (i.e. two elements related in both directions must be the same).

A.4.2 Equivalence Relations

Let A be a set and let R be a relation on A.

- We say R is an **equivalence relation** if and only if R is reflexive, symmetric, and transitive.

- If R is an equivalence relation and $x \in A$, then the **equivalence class corresponding to x (under the relation R)** is

$$[x]_R = \{y \in A \mid (x, y) \in R\}$$

  which is the set of all elements related to x.

- If R is an equivalence relation, then A/R is A **modulo** R; it is the set of all equivalence classes:

$$A/R = \{[x]_R \mid x \in A\}$$

- **Theorem:** If R is an equivalence relation on A, then the equivalence classes (i.e. the elements of A/R) form a *partition* of A.

- **Theorem:** If I is some index set and $\{S_i \mid i \in I\}$ is a *partition* of A, then this corresponds to a unique *equivalence relation* on A defined by relating two elements of A if and only if they belong to the same *part* of the partition.

A.4.3 Modular Arithmetic

Congruence mod n

- Let $n \in \mathbb{N}$ be given. For any $x, y \in \mathbb{Z}$, we say x and y are **congruent modulo** n if and only if $n \mid x - y$.

  Equivalently, this means that x and y have the same remainder upon division by n. (This equivalence is not part of the *definition*; rather, it follows from the Division Lemma stated below.)

  We write this as $x \equiv y \mod n$.

  (Note: mod n is not an *operator* or *function*; it is a *modifier* we place at the end of a line of arithmetic/algebra to indicate that all the operations have been performed modulo n.)

- The relation $\equiv$ is an *equivalence relation*, for every $n \in \mathbb{N}$.

- **Division Lemma:** Let $n \in \mathbb{N}$ be given. Let $x \in \mathbb{Z}$ be given. Then

  $$\exists\,!\, k, r \in \mathbb{Z}\textbf{.}\ \big[(x = kn + r)\ \wedge\ (0 \leq r \leq n - 1)\big]$$

  Notice that "$\exists\,!$" indicates this representation of x as a multiple of n plus a remainder is *unique*.

- **Modular Arithmetic Lemma:** Let $n \in \mathbb{N}$ be given. Let $a, b \in \mathbb{Z}$ be given.

  Suppose that $a \equiv r \mod n$ and $b \equiv s \mod n$. Then,

  $$a + b \equiv r + s \mod n \qquad \text{and} \qquad a \cdot b \equiv r \cdot s \mod n$$

Multiplicative Inverses in $\mathbb{Z}$ mod n

- Let $x, y \in \mathbb{Z}$ be given. We say x and y are **relatively prime** if and only if they have no common factors (divisors), other than 1.

- **MIRP Lemma:** (Multiplicative Inverses for Relative Primes)

  Suppose $n \in \mathbb{N}$ and $a \in \mathbb{Z}$, and that a and n are relatively prime. Consider the congruence $ax \equiv 1 \mod n$. Then there exists a solution x to this congruence.

  (In fact, there are infinitely-many solutions to this congruence, and they are all congruent modulo n.)

- When $ax \equiv 1 \mod n$, we say x is the **multiplicative inverse** of a in the context of $\mathbb{Z}$ modulo n. We write this as $x \equiv a^{-1} \mod n$.

  In fact, any integer y congruent to x modulo n will satisfy $ay \equiv 1 \mod n$, so we really consider the equivalence class $[x]_{\mod n}$ to be the multiplicative inverse of the equivalence class $[a]_{\mod n}$.

- Assuming a^{-1} exists in the first place, $(a^{-1})^{-1} \equiv a \bmod n$.
- Let p be a prime. Then all of the numbers $1, 2, 3, \ldots, p-1$ are guaranteed to be relatively prime to p, so they all have multiplicative inverses in the context of $\mathbb{Z}$ modulo p.
- If a has a multiplicative inverse in the context of $\mathbb{Z}$ mod n, there is guaranteed to be such an inverse between 1 and $n-1$. In practice, we can just check each of these candidates one-by-one until we find the inverse.

Results

- **Chinese Remainder Theorem:** Suppose $r \in \mathbb{N}$ and we have r natural numbers, $n_1, n_2, \ldots, n_r$, that are pair-wise relatively prime. (That is, no two of the numbers have any common factors, besides 1.)

  Suppose we also have r integers, $a_1, a_2, \ldots, a_r$.

  Then there exists a solution X to the system of congruences defined by the n_i and a_i; that is,

  $$\exists X \in \mathbb{Z} \centerdot \forall i \in [r] \centerdot X \equiv a_i \mod n_i$$

  Furthermore, if we define $N = \prod_{i \in [r]} n_i$, then all of the infinitely-many solutions Y to the system of congruences satisfy $X \equiv Y \bmod N$.

A.5 Functions

- Let A, B be sets. Let f be a *relation* between A and B, so $f \subseteq A \times B$.

  Also, assume that f has the property that

  $$\forall a \in A \text{.}\ \exists!\ b \in B \text{.}\ (a, b) \in f$$

  That is, assume every element of the domain A (the "input" set) has *exactly one* corresponding element of the codomain B (the "output" set) so that the two elements are related, under f.

  Put even another way, "Every input has exactly one corresponding output."

  Such a relation is called a **function** from A to B.

  We call A the **domain** of the function and B the **codomain** of the function. We write

  $$f : A \to B$$

  to mean f is a function **from** A **to** B.

  If a is related to b, i.e. $(a, b) \in f$, then we write

  $$f(a) = b$$

  knowing that there is *exactly one* such b for each a.

- Given a proposed domain set A, a proposed codomain set B, and a proposed "rule" f that says what to output, given an element of A, then we say f is a **well-defined function** if the rule is defined on *all* elements of A and, for each $a \in A$, the rule outputs *exactly one* element that actually does lie in the set B.

  (Note: every function is a well-defined function; this rule applies when we are trying to determine whether a given "rule" actually is a function or not.)

- Let $f : A \to B$ and $g : A \to B$ be functions. We say f and g are **equal** (in the sense of functions) and write $f = g$ when $\forall a \in A \text{.}\ f(a) = g(a)$. That is, $f = g$ when the two functions yield the same output for every input.

A.5.1 Images and Pre-Images

- Let $f : A \to B$ be a function. Let $X \subseteq A$. The **image** of X under the function f is

  $$\text{Im}_f(X) = \{b \in B \mid \exists a \in X \text{.}\ f(a) = b\}$$

  An equivalent way of writing this set is

  $$\text{Im}_f(X) = \{f(a) \mid a \in X\}$$

(Intuitively, this is the set of all codomain elements "hit" by elements of X.)

- Let $f : A \to B$ be a function. Let $Z \subseteq B$. The **pre-image** of Z under the function f is
$$\text{PreIm}_f(Z) = \{a \in A \mid f(a) \in Z\}$$
(Intuitively, this is the set of all "inputs" whose output "lands" in Z.)
- Note: $\text{Im}_f(\varnothing) = \varnothing$ and $\text{PreIm}_f(\varnothing) = \varnothing$.

A.5.2 Jections

- Let $f : A \to B$ be a function. If $\text{Im}_f(A) = B$, then we say f is **surjective**, or it is a **surjection**.

  The definition of image gives us this equivalent formulation of surjectivity:
$$f \text{ is surjective } \iff \forall b \in B\,.\, \exists a \in A\,.\, f(a) = b$$
(Intuitively, f is surjective when *all* of the codomain elements are "hit" by the function.)
- Let $f : A \to B$ be a function. If f has the property that
$$\forall a_1, a_2 \in A.\ a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$$
then we say f is **injective**, or it is an **injection**.

  The contrapositive of this conditional statement yields an equivalent formulation of injectivity:
$$\forall a_1, a_2 \in A.\ f(a_1) = f(a_2) \implies a_1 = a_2$$
(Intuitively, f is injective when two different inputs always yield different outputs, or equivalently when having equal outputs means they came from the same input.)
- If a function f is both injective and surjective, then we say f is **bijective**, or it is a **bijection**.

A.5.3 Composition of Functions

- Let $f : A \to B$ and $g : B \to C$ be functions.

  The function $g \circ f : A \to C$ defined by
$$\forall a \in A\,.\, (g \circ f)(a) = g(f(a))$$
is the **composition** of g with f, or "g composed with f".

  Note: It helps to read the "$\circ$" as "after" to remind you of the order of operations: $g \circ f$ means g is applied *after* f. We find $f(a)$ and then find $g(f(a))$.

- Notation: We write $(g \circ f)(x) = g(f(x))$. We do *not* write $g \circ f(x)$. The parentheses are important!
- Let $f : A \to B$ and $g : B \to C$ and $h : C \to D$ be functions. Then $(h \circ g) \circ f = h \circ (g \circ f)$.

  This means **composition is associative**.
- Suppose $f : A \to B$ and $g : B \to C$ are both in/sur/bi-jections. Then $g \circ f$ is also an in/sur/bi-jection.

A.5.4 Inverses

- Let X be any set. The **identity function** $\mathrm{Id}_X : X \to X$ is defined by $\forall z \in X\,.\ \mathrm{Id}_X(z) = z$.
- Let $f : A \to B$ be a function. If there is a function $F : B \to A$ such that $f \circ F : B \to B$ satisfies $f \circ F = \mathrm{Id}_B$ and $F \circ f : A \to A$ satisfies $F \circ f = \mathrm{Id}_A$ then we say F is the *inverse* of f and write $F = f^{-1}$.

  Notice that the formal defintion clearly includes the necessity of checking that *both* ways of composing the two functions yields an identity function. There exist examples where one way works and the other doesn't!

  (Note: When *proving* a function is the inverse of another one, we aren't allowed to write f^{-1} yet because we are, in fact, in the midst of proving that f even *has* an inverse.)

  If f has an inverse, we say f is **invertible**.
- **Theorem:** $f : A \to B$ is bijective $\iff$ f has an inverse $f^{-1} : B \to A$.
- **Theorem:** Let $f : A \to B$ and $g : B \to C$ both be bijections. Then $g \circ f : A \to C$ is also a bijection, so it has an inverse; that inverse is $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

A.5.5 Proof Techniques for Functions

- Prove that f is **surjective**:
  - Let $b \in B$ be arbitrary and fixed.
  - Define $a =$ _____.
  - Show that $a \in A$.
  - Show that $f(a) = b$.
  - This shows that $b \in \mathrm{Im}_f(A)$, so $B \subseteq \mathrm{Im}_f(A)$.
  - Since $\mathrm{Im}_f(A) \subseteq B$ by definition, this shows $\mathrm{Im}_f(A) = B$, so f is surjective.
- Prove that f is **not surjective**:

  - Define $b =$ _____.
  - Show that $b \in B$.
  - Let $a \in A$ be arbitrary and fixed.
  - Show that $f(a) \neq b$. (Alternatively, suppose $f(a) = b$ and find a contradiction.)
  - This shows that $\exists b \in B \mathbin{.}\ b \notin \mathrm{Im}_f(A)$, so f is not surjective.

- Prove that f is **injective**:
  - Let $x, y \in A$ be arbitrary and fixed.
  - Suppose that $f(x) = f(y)$.
  - Deduce that $x = y$.

  Alternatively:
  - Let $x, y \in A$ be arbitrary and fixed.
  - Suppose that $x \neq y$.
  - Deduce that $f(x) \neq f(y)$.

- Prove that f is **not injective**:
  - Define $x =$ _____ and define $y =$ _____.
  - Show that $x \in A$ and $y \in A$.
  - Show that $x \neq y$.
  - Show that $f(x) = f(y)$.
  - This shows $\exists x, y \in A \mathbin{.}\ x \neq y \ \wedge\ f(x) = f(y)$, so f is not injective.

- Prove that f is **bijective**:
  - Prove that f is injective.
  - Prove that f is surjective.

  Alternatively:
  - Define a function $F : B \to A$.
  - Prove that $F \circ f = \mathrm{Id}_A$.
  - Prove that $f \circ F = \mathrm{Id}_B$.
  - This shows that $F = f^{-1}$, so f is invertible and therefore it is bijective.

- Prove that f is **not bijective**:
  - Show that f is not injective, or else show that f is not surjective.

  Alternatively,

  - AFSOC f is bijective, so it has an inverse f^{-1}. Find a contradiction.

- For some $X \subseteq A$, find the image $\text{Im}_f(X)$:
  - Define a set S. Claim that $S = \text{Im}_f(X)$.

    (Note: Coming up with this definition is the hard part, and will involve a bunch of scratch work. There is no need to show this as part of your proof. Just start with the definition.)
  - Prove that $\text{Im}_f(X) \subseteq S$.
    * Let $y \in \text{Im}_f(X)$ be arbitrary and fixed.
    * This means $\exists a \in X \,.\, f(a) = y$.
    * Use the properties of f to show that $f(a) \in S$.
    * This shows that $y \in S$.
  - Prove that $S \subseteq \text{Im}_f(X)$.
    * Let $z \in S$ be arbitrary and fixed.
    * Define $x =$ _____.
    * Show that $x \in X$.
    * Show that $f(x) = z$.
    * This shows that $z \in \text{Im}_f(X)$.
  - Conclude by a double-containment argument that $\text{Im}_f(X) = S$.

- For some $Z \subseteq B$, find the preimage $\text{PreIm}_f(Z)$:
  - Define a set T. Claim that $T = \text{PreIm}_f(Z)$.

    (Note: Coming up with this definition is the hard part, and will involve a bunch of scratch work. There is no need to show this as part of your proof. Just start with the definition.)
  - Prove that $\text{PreIm}_f(Z) \subseteq T$.
    * Let $a \in \text{PreIm}_f(Z)$ be arbitrary and fixed.
    * This means $f(a) \in Z$.
    * Use the properties of f to show that $a \in T$.
  - Prove that $T \subseteq \text{PreIm}_f(Z)$.
    * Let $x \in T$ be arbitrary and fixed.
    * Use the properties of f to show that $f(x) \in Z$.
    * This shows that $x \in \text{PreIm}_f(Z)$.
  - Conclude by a double-containment argument that $\text{PreIm}_f(Z) = T$.

- Find the **inverse** of f.
  - Define a function $F : B \to A$.

    (Note: Coming up with this definition is the hard part, and will involve a bunch of scratch work. There is no need to show this as part of your proof. Just start with the definition.)

- Show that F is a well-defined function: show that every input from B has exactly one output that lies in A.
- Show that $F \circ f = \mathrm{Id}_A$.
- Show that $f \circ F = \mathrm{Id}_B$.
- Deduce that $F = f^{-1}$. (Since f has an inverse, it is therefore a bijection, as well.)

A.6 Cardinality

A.6.1 Definitions

Let S be any set.

- We say S is **finite** if $\exists n \in \mathbb{N} \cup \{0\}$ such that there exists a bijection $f : S \to [n]$.

  Note: The empty set $S = \varnothing$ is finite, since $[0] = \varnothing$.

- We say S is **infinite** if S is *not finite*; that is, if $\forall n \in \mathbb{N} \cup \{0\}$, every function $f : S \to [n]$ fails to be a bijection.

- We say S is **countably infinite** (or just **countable**) if there exists a bijection $f : S \to \mathbb{N}$.

- We say S is **uncountably infinite** (or just **uncountable**) if every function $f : S \to \mathbb{N}$ fails to be a bijection.

- We use $|S|$ to indicate the **cardinality** of S.

  When S is finite, so there is some $n \in \mathbb{N} \cup \{0\}$ and a bijection $f : S \to [n]$, we write $|S| = n$ to mean that S has n elements. We say n is the **size** of S.

  When S is infinite, we only use $|S|$ to **compare** the cardinality of S to that of other sets. That is, we don't write things like $|S| = \infty$; rather, we write something like $|S| = |T|$ to indicate that S and T have the *same* cardinality, whatever that may be, or something like $|S| < |T|$ to indicate T has a *strictly larger* cardinality than S.

- We write $|S| = |T|$ and say S has the **same cardinality** as T if and only if there exists a bijection $f : S \to T$.

A.6.2 Results

In general, the following results hold. Some of the remaining results follow from these general statements.

- Suppose $|A| = |C|$ and $|B| = |D|$. Then $|A \times B| = |C \times D|$.

- Suppose $|A| = |C|$ and $|B| = |D|$, and suppose $A \cap B = \varnothing$ and $C \cap D = \varnothing$. Then $|A \cup B| = |C \cup D|$.

- Suppose there is an injection $f : A \to B$. Then $|A| \leq |B|$.

- Suppose there is a surjection $f : A \to B$. Then $|A| \geq |B|$.

Finite Sets

- If A and B are finite, then $A \cup B$ are finite.
- If A and B are finite and $A \cap B = \varnothing$, then $|A \cup B| = |A| + |B|$.
- If A and B are finite, then $|A \times B| = |A| \cdot |B|$.

Infinite Sets

- If A is countably infinite and B is finite or countably infinite, then $A \cup B$ is countably infinite.
- If A is countably infinite and B is finite or countably infinite, then $A \times B$ is countably infinite.
- If A is uncountably infinite and B is any set, then $A \cup B$ is uncountably infinite.
- If A is uncountably infinite and B is any set, then $A \times B$ is uncountably infinite.
- If $A \subseteq B$, then $|A| \leq |B|$. (Note: this applies to both finite and infinite sets.)
- $|A| < |\mathcal{P}(A)|$ for any set A. (Note: this applies to both finite and infinite sets!)
- If A is infinite, then there exists a set $C \subseteq A$ that is countably infinite.
- A is infinite $\iff \exists C \subset A$ such that there exists a bijection $f : A \to C$. (Note the *strict* subset inequality.)
- A countably infinite union of countably infinite sets is also countably infinite.
- A countably infinite product of finite sets is uncountably infinite.

  (Notice that this shows that a countably infinite product of any non-empty sets is uncountably infinite.)
- **Cantor-Schröder-Bernstein Theorem:** Suppose A and B are sets and there exist functions $f : A \to B$ and $g : B \to A$ that are both *injections*. Then there actually exists a *bijection* $h : A \to B$ so, in particular, $|A| = |B|$.

A.6.3 Standard Catalog of Cardinalities

- **Finite sets:**
  - $\varnothing$
  - $[n]$, for any $n \in \mathbb{N}$
- **Countably Infinite sets:**
  - $\mathbb{N}$
  - $\mathbb{Z}$
  - Odd naturals/integers, Even naturals/integers
  - $\mathbb{Q}$
  - $\mathbb{N} \times \mathbb{N}$
  - The set of all *finite* binary strings
- **Uncountably Infinite sets:**
  - $\mathbb{R}$
  - Intervals of $\mathbb{R}$; that is, $\{y \in \mathbb{R} \mid a \leq y \leq b\}$ for any $a, b \in \mathbb{R}$.

    (Note: the "$\leq$" in the interval can each be replaced by "$<$" as well.)
  - $\mathcal{P}(\mathbb{N})$
  - $\mathcal{P}(\mathbb{Z})$
  - The set of all *infinite* binary strings

A.7 Combinatorics

A.7.1 Definitions

- A **permutation** of the set $[n]$ is a bijection $f : [n] \to [n]$.
- A **k-selection** from the set $[n]$ is a subset $S \subseteq [n]$ with $|S| = k$.
- A **k-arrangement** from the set $[n]$ is an ordered list of k elements of $[n]$, where no element is repeated.
- A **k-selection with repetition** from the set $[n]$ is an unordered list of k elements of $[n]$, where elements can repeat.
- A **k-arrangement with repetition** from the set $[n]$ is an ordered list of k elements of $[n]$, where elements can repeat.

A.7.2 Counting Principles

- **Rule Of Sum:** Let A be a finite set. Let $n \in \mathbb{N}$. Suppose $\{S_i \mid i \in [n]\}$ is a partition of A. Then

$$|A| = \sum_{i=1}^{n} |S_i| = |S_1| + |S_2| + \cdots + |S_n|$$

- **Rule Of Product:** Suppose we have a process that is completed in n steps. Suppose that step i (where $1 \leq i \leq n$) can be completed in w_i ways, independent of the choices made in the previous step. Then the number of outcomes of this process is

$$\prod_{i=1}^{n} w_i = w_1 \cdot w_2 \cdots w_n$$

A.7.3 Formulas

- There are $n!$ many permutations of $[n]$.
- There are $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ many k-selections from $[n]$.
- There are $\binom{n}{k}\, k! = \frac{n!}{(n-k)!}$ many k-arrangements from $[n]$.
- There are $\binom{k+n-1}{k}$ many k-selections with repetition from $[n]$.
- There are n^k many k-arrangements with repetition from $[n]$.

A.7.4 Standard Counting Objects

- **Cards:** A standard deck of cards has 52 cards. Each card has a suit (either ♡ or ♢ or ♣ or ♠) and a rank (either 2 or 3 or 4 or ... or 10 or Jack or Queen or King or Ace).
- **Tuples:** Let $k, n \in \mathbb{N}$. The set $T_{n,k}$ is the set of all n-tuples from $[k]$. That is, it is the set of all ordered lists of length n whose coordinates are elements of $[k]$.
- **Words:** This is equivalent to tuples, where $[k]$ represents the alphabet and n represents the length of a word.
- **Lattice Paths:** Let $x, y \in \mathbb{N}$. A lattice path to (x, y) is a sequence of points on the grid of natural-numbered points on the plane, starting at $(0, 0)$ and ending at (x, y), where each successive move is either rightwards or upwards.

  There are $\binom{x+y}{x} = \binom{x+y}{y}$ many lattice paths to (x, y).

A.7.5 Counting In Two Ways

This is a standard method for proving an identity using a combinatorial argument.

Method Outline:

1. State the result to be proven. Note: remember to quantify any variables that appear in the expression!
2. Define a set (let's call it S) of objects to be counted.
3. Count the elements of S in one way by following a proper combinatorial argument. Equate the derived expression with $|S|$.
4. Count the elements of S in another way by following a proper combinatorial argument. Equate the derived expression with $|S|$.
5. Conclude that since both derived expressions equal $|S|$, they must be equal, as well.

A.7.6 Results

These were proven in lecture by counting in two ways arguments. (You may cite these results without proof, but it is also helpful to remember the main idea of the counting arguments, as well, so that you can reconstruct the formula without having to just memorize it.)

- **Pascal's Identity:** $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

- **Chairperson Identity:** $\binom{n}{k} \cdot k = n \cdot \binom{n-1}{k-1}$
- **Binomial Theorem:** $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$
- **Summation Identity:** $\sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$

A.7.7 Inclusion/Exclusion

We have a universal set U and some subsets $A_1, A_2, \ldots, A_n \subseteq U$. We want to count the elements of U that are *outside* of all of the A_i sets.

$$\begin{aligned} |U - A_1| &= |U| - |A_1| \\ |U - (A_1 \cup A_2)| &= |U| - |A_1| - |A_2| + |A_1 \cap A_2| \\ |U - (A_1 \cup A_2 \cup A_3)| &= |U| - |A_1| - |A_2| - |A_3| \\ &\quad + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| \\ &\quad - |A_1 \cap A_2 \cap A_3| \end{aligned}$$

and so on.

In general, for n many sets, we have

$$|U - (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{S \subseteq [n]} (-1)^{|S|} \left| \bigcap_{i \in S} A_i \right| \qquad \text{where} \quad \bigcap_{i \in \varnothing} A_i = U$$

In the (convenient) case where the *size* of the intersection of k-many sets only depends on that value k (and not *which* sets we are intersecting), then we can write

$$|U - (A_1 \cup A_2 \cup \cdots \cup A_n)| = \sum_{k=0}^{n} (-1)^k \binom{n}{k} |S_1 \cap S_2 \cap \cdots \cap S_k|$$

A.7.8 Pigeonhole Principle

If a set S with $|S| = n$ is partitioned into k disjoint subsets whose union is S, and if $k < n$, then at least one of the subsets in the partition has more than one element. Furthermore, that part actually has at least $\lceil \frac{n}{k} \rceil$ elements.

(That is, if we separate n objects into k piles, there must be one pile with at least $\frac{n}{k}$ objects in it.)

A.8 Acronyms

A.8.1 General Phrases

- **WWTS**: We want to show
- **AFSOC**: Assume for sake of contradiction
- **WOLOG**: Without loss of generality

A.8.2 Induction

- **PMI**: Principle of Mathematical Induction
- **BC**: Base case
- **IH**: Induction hypothesis
- **IS**: Induction step